Fall 2013

Manifold learning based spectral unmixing of hyperspectral remote sensing data

Jun-Hwa Chi
Purdue University

MANIFOLD LEARNING BASED SPECTRAL UNMIXING OF HYPERSPECTRAL REMOTE SENSING DATA

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Jun-Hwa Chi

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2013

Purdue University

West Lafayette, Indiana

ACKNOWLEDGMENTS

I would like to gratefully and sincerely thank my major advisor, Professor Melba Crawford for her guidance, understanding and patience. This dissertation would have not been finished without her insightful comments on my research directions and providing excellent atmosphere for doing research as well as financial support during my PhD program. I would also like to express my deepest gratitude to all of the committee members: Professor James Bethel, Professor Okan Ersoy and Professor Jie Shan for helping me to develop my academic and research background.

My appreciation also goes to my colleagues at the Laboratory for Applications of Remote Sensing, Wonkook, Jinha, Lexie, Garett, Darion and Magda. I would never forget academic discussion and field campaigns with them. I also enjoyed every moment in West Lafayette with Geomatics graduate students and my friends in PKCE and PKTC.

Finally and most importantly, I would like to thank my parents whose endless love, enormous sacrifices, constant encouragement and confidence in me made this dissertation possible.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

ABBREVIATIONS

DR	Dimensionality Reduction
SNR	Signal-to-Noise Ration
EE	Endmember Extraction
PCA	Principal Component Analysis
MNF	Maximum (or Minimum) Noise Fraction
ISOMAP	Isometric feature mapping
L-ISOMAP	Landmark Isometric feature mapping
LLE	Locally Linear Embedding
MDS	Multidimensional Scaling
LH	Local Homogeneity
HM	Hybrid Manifold
AVIRIS	Airborne Visual and Infra-Red Imaging Spectrometer
SAD	Spectral Angle Distance
AL	Active Learning
RMSE	Root Mean Square Error
SSV	Sequential Simplex Volume
SVD	Singular Value Decomposition
MaxDist/MinDist	Maximum/Minimum spectral distance
CAI	Cellulose Absorption Index
ACRE	Agronomy Center for Research and Education
SWIR	Short Wave Infrared

ABSTRACT

Chi, Jun-Hwa. Ph.D., Purdue University, December 2013. Manifold Learning Based Spectral Unmixing of Hyperspectral Remote Sensing Data. Major Professor: Melba M. Crawford.

Nonlinear mixing effects inherent in hyperspectral data are not properly represented in linear spectral unmixing models. Although direct nonlinear unmixing models provide capability to capture nonlinear phenomena, they are difficult to formulate and the results are not always generalizable. Manifold learning based spectral unmixing accommodates nonlinearity in the data in the feature extraction stage followed by linear mixing, thereby incorporating some characteristics of nonlinearity while retaining advantages of linear unmixing approaches. Since endmember selection is critical to successful spectral unmixing, it is important to select proper endmembers from the manifold space. However, excessive computational burden hinders development of manifolds for large-scale remote sensing datasets. This dissertation addresses issues related to high computational overhead requirements of manifold learning for developing representative manifolds for the spectral unmixing task.

Manifold approximations using landmarks are popular for mitigating the computational complexity of manifold learning. A new computationally effective landmark selection method that exploits spatial redundancy in the imagery is proposed. A robust, less costly landmark set with low spectral and spatial redundancy is successfully incorporated with a hybrid manifold which shares properties of both global and local manifolds.

While landmark methods reduce computational demand, the resulting manifolds may not represent subtle features of the manifold adequately. Active learning heuristics are introduced to increase the number of landmarks, with the goal of developing more representative manifolds for spectral unmixing. By communicating between the landmark set and the query criteria relative to spectral unmixing, more representative and stable manifolds with less spectrally and spatially redundant landmarks are developed. A new ranking method based on the pixels with locally high spectral variability within image subsets and convex-geometry finds a solution more quickly and precisely. Experiments were conducted to evaluate the proposed methods using the AVIRIS Cuprite hyperspectral reference dataset.

A case study of manifold learning based spectral unmixing in agricultural areas is included in the dissertation.Remotely sensed data collected by airborne or spaceborne sensors are utilized to quantify crop residue cover over an extensive area. Although remote sensing indices are popular for characterizing residue amounts, they are not effective with noisy Hyperion data because the effect of residual striping artifacts is amplified in ratios involving band differences. In this case study, spectral unmixing techniques are investigated for estimating crop residue as an alternative approach to empirical models developed using band based indices. The spectral unmixing techniques, and especially the manifold learning approaches, provide more robust, lower RMSE estimates for crop residue cover than the hyperspectral index based method for Hyperion data.

1. INTRODUCTION

1.1 Hyperspectral Remote Sensing

Recent advances in remote sensing, motivated by a desire to detect and characterize detailed information about targets of interest, have led to the development of advanced sensors that acquire more detailed spectral signatures. Hyperspectral remote sensing, also known as imaging spectroscopy, is a relatively new technology which was developed in the laboratory by physicists and chemists over a period of 100 years to detect individual absorption features related to specific chemical bonds in a solid, liquid, or gas [1–3].

Hyperspectral remote sensing has gained increased attention in many remote sensing applications which have focused on the Earth during the past decade because of the capability to record reflected radiation from a ground target over a continuous range of contiguous bands. The enhanced capability of high spectral resolution sensors, which often acquire data in at least 100 spectral bands with narrow bandwidths (5 - 10 nm), provides improved capacity for identifying materials compared to that of multispectral sensors [4].

In recent years, airborne hyperspectral sensing has become relatively common, and the NASA (National Aeronautics and Space Administration) Hyperion sensor on Earth Observing-1 (EO-1) has provided the foundation for upcoming spaceborne missions such as the JAXA (Japan Aerospace Exploration Agency) HISUI (Hyperspectral Imager Suite), ESA (European Space Agency) EnMAP (Environmental Mapping and Analysis Program), and NASA HyspIRI. The methods developed in this dissertation should facilitate utilization of data from these missions.

1.2 Challenges for Hyperspectral Remote Sensing

Because high spectral resolution data allow improved capability for discrimination of subtle differences in ground objects, which is not possible with multispectral data, hyperspectral data analysis has become an active area of research. However, the image analysis techniques used in multispectral data cannot simply be extended to hyperspectral image processing because of significant challenges existing in high dimensional data.

1.2.1 High Inter-band Correlation

Highly correlated narrow, contiguous spectral bands for a given hyperspectral pixel are usually problematic both for processing and analysis. In classification tasks, for example, higher dimensional feature spaces do not always guarantee better classification accuracy with a limited number of samples, which was demonstrated by Hughes [5–7]. This implies that classification accuracy decreases when the number of input features exceeds the optimal value for a given dimensionality. Moreover, it is difficult to estimate parameters for statistically based techniques, including classification and feature extraction methods, since the covariance matrix of highly correlated data is near singular. Recently, airborne hyperspectral sensors, which exploit the advantages of both spatial and spectral resolution, have increased the quantity of high dimensional data, and several spaceborne hyperspectral missions have been planned. As the capability of the sensors improves and the availability of data increases, methodological advances are needed to analyze these data.

1.2.2 Nonlinearity in Hyperspectral Remote Sensing Data

The data points of hyperspectral imagery typically lie on a subspace of much lower dimensionality. Additionally, the nonlinear phenomena often exhibited in hyperspectral imagery (see Fig. 1.1) are problematic for analysis of hyperspectral data

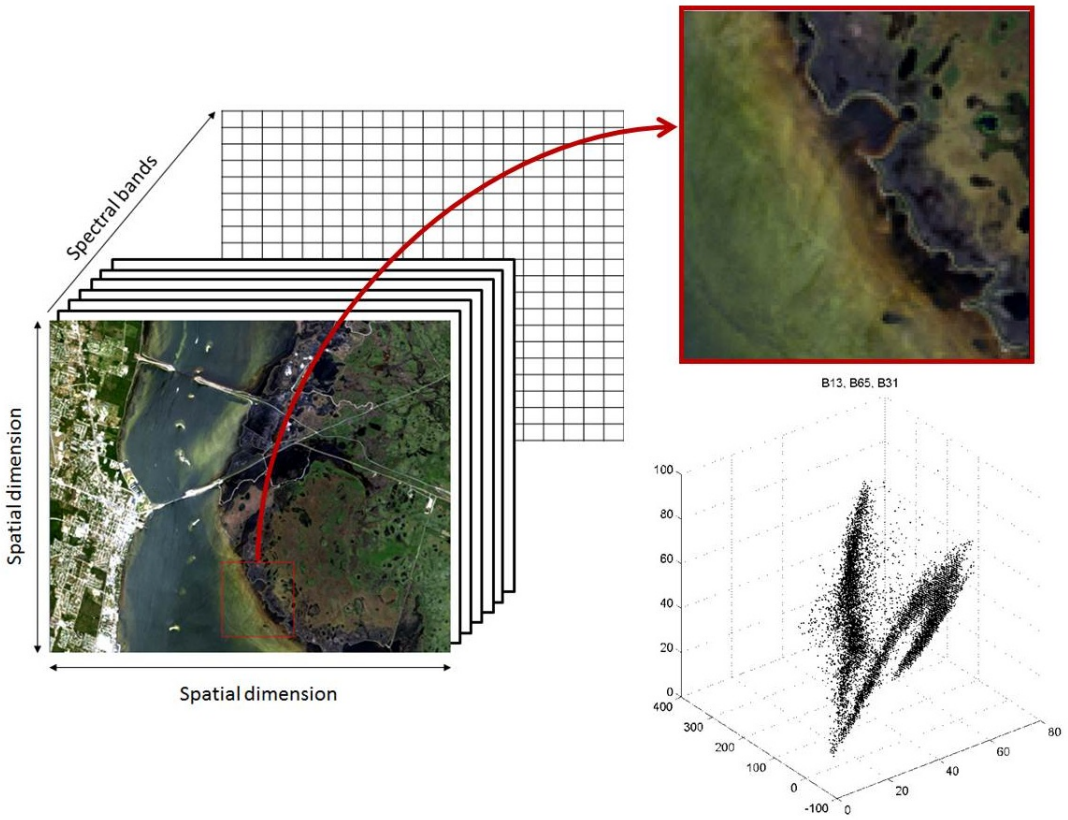


Figure 1.1. Nonlinearity inherent in hyperspectral remote sensing data

and development of associated derived products [8]. Nonlinearity in hyperspectral data can be attributed to various sources, including variations in the bi-directional reflectance distribution function (BRDF) due to surface and atmospheric properties [9, 10], multi-path scattering between heterogeneous pixel constituents [11], and nonlinear attenuation [8, 12] in vegetation and water bodies. Linear signal processing methods do not address the nonlinearity in hyperspectral data, often resulting in non-representative features and low classification and unmixing accuracies.

1.3 Spectral Mixture Analysis

In medium spatial resolution hyperspectral imagery, which is common for satellite based sensors, most pixels are a mixture of more than one spectral signature for several reasons:

- There is a tradeoff between spatial and spectral resolution of a sensor since the energy detected by a sensor represents the integrated response over a range of wavelengths and is associated with ground resolution cell of given spatial extent. Thus, although hyperspectral sensors can record rich spectral information over narrow bands, thereby providing improved discrimination capability, the spatial resolution is often lower than that provided by corresponding multispectral sensors. Airborne hyperspectral sensors acquire images at a lower altitude with higher spatial resolution, but at a much higher cost and lower acquisition frequency. Since the spatial resolution of typical space-based hyperspectral data is not adequate for separating pure signatures, most pixels contain signatures of multiple materials [13]. Further, the coupling between spatial and spectral resolution results optical aberrations of the two domains in the instrument point spread function (PSF) [14].

- In geological images, mixtures of spectra associated with different classes may exist mixtures at a microscopic scale, also referred to as intimate mixtures [15].

- The most difficult limitation is attributed to the effects of multiple interference contributed by atmospheric attenuation, varying illumination conditions, multiple scattering and shadow effects which are difficult to model and remove [16].

Thus, increasing the spatial resolution does not fully solve the mixed pixel issue. Therefore, in the last decade, spectral mixture analysis has received significant attention and has been widely studied for hyperspectral data analysis and quantification of the spatial coverage of given land cover classes and materials images [17].

1.4 Spectral Unmixing Approaches

Spectral unmixing decomposes a mixed pixel into a collection of individual pure spectral signatures at sub-pixel levels [13, 17–23]. Unmixing methods are usually comprised of two phases:

1. Finding the spectrally pure or extreme signatures, referred to as "*endmembers*," that can be used to "unmix" remaining mixed pixels in the data.

2. Expressing each pixel in the image as linear or nonlinear combinations of the endmembers, and computing the corresponding fractional abundances which are related to "physical" amounts of the endmembers.

Endmembers normally correspond to familiar spectrally homogeneous objects in the scene, such as water, soil and vegetation [13, 17–19].

1.4.1 Linear Spectral Unmixing

Linear mixture models solve the mixed pixel problem based on the assumption that reflectances of the individual pure materials are linearly independent, and the pixels in the image lie in linear spaces. This can be described as follows:

Given data samples in a hyperspectral data matrix $[\mathbf{X} : \mathbf{x}_i, i = 1, \ldots, N] \in \mathbf{R}^D$, where D is the original dimension and N is the number of samples, let $\mathbf{x} = (s, t)$ be the hyperspectral signature collected by the sensor at the pixel with spatial coordinates (s, t). This spectrum can be considered as a D-dimensional vector and might be modeled in terms of a linear combination of several endmember vectors $[\mathbf{E} : \mathbf{e}_i, i = 1, \ldots, q]$, where q is the number of endmembers, using the following expression:

$$\mathbf{x}\,(s, t) = \sum_{i=1}^{q} a_i\,(s, t)\,\mathbf{e}_i \tag{1.1}$$

where $a_i(s,t)$ is a scalar value representing the fractional coverage of endmember vector $\mathbf{e}_i$ at pixel $\mathbf{x}(s,t)$ [13, 17–19]. In order to estimate physically meaningful abundances, two constraints are usually imposed: 1) the nonnegativity (1.2) and 2) sum-to-one constraints (1.3), respectively, defined as [24]

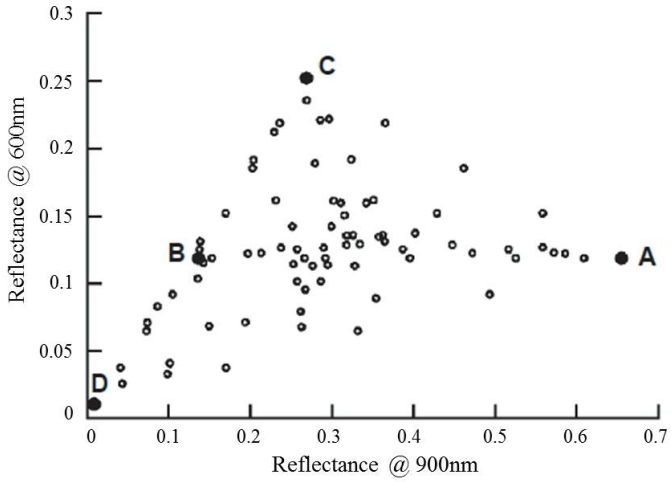
$$a_i(s,t) \geq 0, \qquad \forall a_i : 1 \leq i \leq q \tag{1.2}$$

$$\sum_{i=1}^{q} a_i(s,t) = 1 \tag{1.3}$$

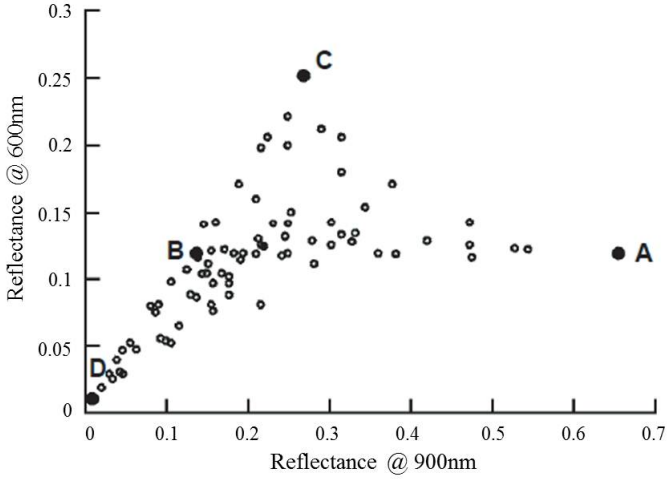
This is referred to as fully constrained unmixing.

1.4.2 Nonlinear Spectral Unmixing

Linear unmixing is often employed because of the ease of implementation, robustness and the difficulty in developing appropriate nonlinear models. However, linear mixing models might not be adequate for representing the mixing phenomena in nature. Nonlinear effects between different surface components, which violate the linearity assumption of mixing models [25], have been recognized for many years in analysis of mineral mixtures, as well as for vegetation and canopy studies [13, 25, 26]. Fig. 1.2 illustrates the effects of nonlinear mixing on reflectance using four endmembers. While the mixture points in two spectral bands shown in Fig. 1.2(a) are spread along the lines connected with the endmembers, Fig. 1.2(b) shows that the point cloud joining the endmembers is curvilinear, which indicates the nonlinear effects clearly [13]. Utilization of linear unmixing models with data that contain nonlinear effects results in inaccurate extracted endmembers and computed abundance fractions. In [26], Herzog and Mustard calculated fractions between linear and nonlinear mixing in laboratory data. The resulting errors between the two models were as much as 30%.



(a)



(b)

Figure 1.2. Scatter plots with mixing models: (a) Linear mixing and (b) Nonlinear mixing

Currently, two general strategies are employed to deal with nonlinear mixing effects in hyperspectral data: 1) direct nonlinear unmixing models, and 2) capturing nonlinearity in the dimensional reduction (DR) stage.

Direct nonlinear unmixing techniques have yet to adequately represent the physical phenomena due to the complexity of these solutions. A few studies focused on direct nonlinear modeling via kernel based and neural network models have been conducted [27–31]. Although direct nonlinear unmixing models provide more capability to capture nonlinear mixing effects than the linear models, they are difficult to formulate, analyze, and derive simple mathematical solutions, and may not be applicable to a broad range of datasets [27].

Over the past decade, nonlinear feature extraction methods, also known as "*manifold learning*," have been proposed by the machine learning community to address nonlinearity exhibited in data. These methods have been used recently to address nonlinearity in hyperspectral remote sensing data analysis [32–39]. Although nonlinear manifolds have been investigated extensively for classification [32–36], only limited research has been undertaken on spectral unmixing [37–39]. Recently, manifold learning approaches, which address nonlinearity in the DR stage, have received attention as an alternative approach to the direct nonlinear unmixing models [37–39].

1.5 Organization of the Dissertation

In this dissertation, manifold learning-based spectral unmixing is exploited to accommodate nonlinearity in hyperspectral remotely sensed data. Chapter 2 summarizes a brief background and related work in spectral unmixing. Chapters 3 and 4 present new methodologies to develop cost effective, representative manifolds for unmixing. In Chapter 5, an agricultural application of unmixing is investigated as an alternative to a traditional index based method. A brief overview of the dissertation is outlined in Fig. 1.3 and as follows:

- **Chapter 2. Background and Related work in Spectral Unmixing**

  This chapter reviews popular endmember extraction algorithms, which are the most critical part of spectral unmixing, as well as linear feature extraction and manifold learning methods which are key to this study. Approximation-based manifold learning approaches to reduce computational complexity are also summarized. Additionally, some of the major contributions in both direct nonlinear unmixing models and manifold learning based unmixing approaches are summarized in this chapter.

- **Chapter 3. Landmark Based Manifold Learning for Nonlinear Unmixing**

  High computational overhead of manifold learning is problematic for large scale remotely sensed data. A small set of distinguishing points, referred to as "*landmarks*," is widely used to reduce the computational cost. In this chapter, a landmark selection method is proposed to exploit both spectral and spatial proximity of the data. Moreover, a hybrid manifold, which shares character-

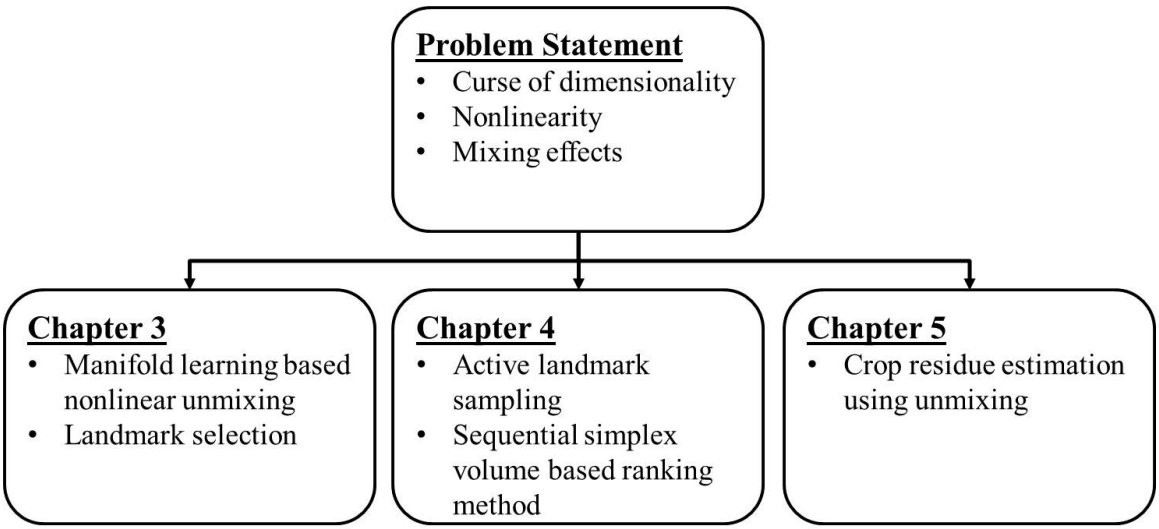


Figure 1.3. Overview of the dissertation

istics of both global and local manifolds, is proposed to further reduce computational complexity. The proposed methods are compared qualitatively and quantitatively.

- **Chapter 4. Active Landmark Sampling in Manifold Learning Based Unmixing**

  The proposed landmark selection method in Chapter 3 automatically determines a fixed number of resulting landmarks. When manifolds have complex geometry, additional landmarks may be needed to characterize the manifold smoothly and to increase its stability. This chapter provides an "*active sampling strategy*," which is motivated by active learning in supervised classification to improve the representation of the manifold. A new ranking method using sequential image subsets to determine candidates is demonstrated and compared to spectral distance based ranking methods.

- **Chapter 5. Crop Residue Estimation Using Unmixing Techniques**

  In this chapter, a spectral unmixing application in agricultural areas is demonstrated. A remote sensing index-based method is commonly used to estimate the amount of crop residue. However, this approach is problematic for low SNR spaceborne hyperspectral imagery due to noise that is exacerbated in the differences and ratios, and ground measurements are required to develop empirical models for estimating residue. Spectral unmixing-based techniques can mitigate the impact of these problems. Qualitative and quantitative experiments are conducted to compare the index-based method to linear and nonlinear manifold learning-based unmixing techniques.

- **Chapter 6. Conclusions and Future Work** The primary contributions of the dissertation are summarized and discussed. Open problems for future work are also considered.

2. BACKGROUND AND RELATED WORK IN SPECTRAL UNMIXING

2.1 Endmember Extraction Algorithms (EEAs)

The most critical task in linear mixture analysis is identification of an appropriate set of endmembers to use in modeling at-sensor pixel spectra through a combination of the endmembers. The selection of endmembers can be accomplished directly from the image, from the field, or from library spectra of well-known targets. The risk in using field or library spectra is that the spectra are rarely acquired under the same conditions as the remote sensing data. Image derived endmembers also have the advantage of being collected on the same scale as the image data and can be more easily associated with features on the scene [13]. The categories of endmember identification approaches shown in Fig. 2.1 represent algorithms that have been proposed over the past decade to accomplish the task of finding appropriate image derived endmembers for spectral mixture analysis.

2.1.1 Spectral Based Endmember Extraction Algorithms

Traditional, widely used endmember extraction methods only exploit spectral information from the image. Four main strategies have been used to design spectral-based EEAs: 1) principal orthogonal projection (e.g., PPI), 2) use of convex geometry as a criterion (e.g., CCA, N-FINDR), 3) linear spectral mixture analysis (e.g., IEA), and 4) second-order statistics (e.g., SPCA-EEA). Following is a brief overview of popular traditional EE methods:

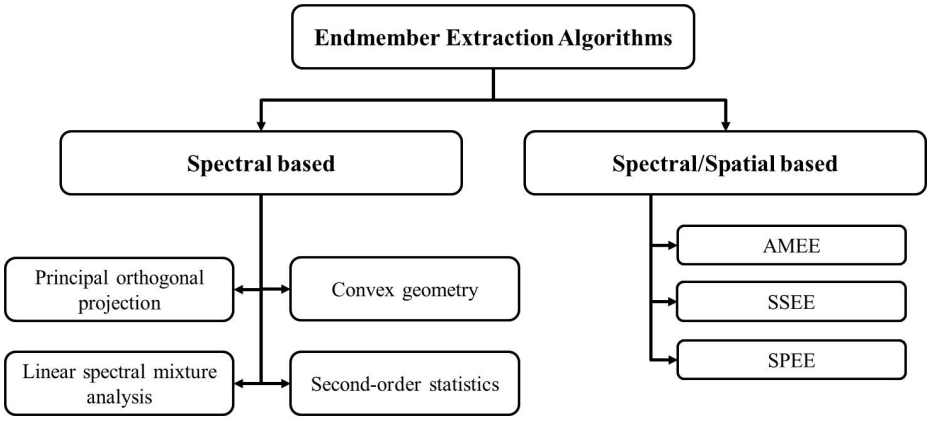


Figure 2.1. Categories of EEAs

1. *Principal orthogonal projection*

   - Pixel purity index (PPI)

     Selection by the PPI [22] is accomplished by randomly generated lines, referred to as skewers, in the d-dimensional space comprised of a scatter plot of the maximum noise fraction (MNF) [40] transformed space. All the points in the space are projected onto the lines, and those pixels falling at the extremes of the lines are counted. After repeated projections onto different lines, pixels with a count above a threshold are declared "pure". Since there are many redundant pure spectra, a final endmember is selected through d-dimensional visual analysis. Although the PPI is a popular and widely used technique, the sensitivity of the two parameters, the number of skewers and the cut-off threshold, is a major issue. Manual selection of the final set of endmembers by human intervention is also a limitation of the PPI.

2. *Convex geometry*

   Assuming that pure signatures are endmembers, the concept of convexity of geometry is natural and logical. Thus, this general approach is the most popular, resulting in development of a wide range of algorithms.

   - Convex cone analysis (CCA)

     In the CCA [41] algorithm, physical quantities (radiance or reflectance) are assumed to be nonnegative. The vectors formed by radiance/reflectance spectra can be expressed as linear combinations of nonnegative components which lie at the extremes of a nonnegative, convex region. The goal of this algorithm is to find the extreme points for the convex region. The method finds the eigenvectors of the sample spectral correlation matrix of the image, and selects those eigenvectors corresponding to the largest eigenvalues. It then determines the boundaries of the convex cone, where there are strictly nonnegative linear combinations of these eigenvectors.

   - N-FINDR

     The N-FINDR [42] algorithm is an automated technique for finding endmembers in the image. The method attempts to find the simplex of maximum volume spanned by the endmembers. It is based on the assumption that vertices of the simplex which yield the largest volume will likely comprise the purest pixels. If hyperspectral data are distributed over a convex space, this operation can be performed in a quick and relatively straightforward manner. The N-FINDR method was extended to improve the performance through methods such as single/multiple replacement, successive iterative N-FINDR, and the simplex growing algorithm [43–47].

3. *Linear spectral mixture analysis (LSMA)*

   - Iterative error analysis (IEA)

     The IEA [48] method identifies pixel spectra as image endmembers iteratively. In each iteration, fully constrained spectral unmixing based on previously identified endmembers is performed to model all the pixels in the image. The pixel spectrum with the largest residual error is then selected as a new image endmember.

   - Iterated constrained endmember (ICE)

     Berman et al. [49] proposed the ICE, which minimizes the unmixed error from a simplex spanned by q endmembers. While convex geometry-based EEAs such N-FINDR use the simplex volume, the ICE constrains the size of the simplex via the sum of Euclidean distance among all selected endmembers.

4. *Second-order statistics*

   - Standardized principal components analysis based EEA (SPCA-EEA)

     Singh and Harison who expressed the second-order statistics information in terms of correlation coefficients explored the second-order statistical correlation [50]. Their idea motivated to develop the SPCA by Eklundh and Singh [51]. In [52] , author assumed that the information represented by the second-order statistics among endmembers is minimal. The spectral correlation coefficients among the q endmembers is the smallest that represented by all possible q signatures. If one of q endmembers is a mixture of other (q-1) distinctive signatures, the shared second-order statistics must be at least equal to or greater than those represented by the q endmembers.

2.1.2 Spectral and Spatial Based Endmember Extraction Algorithms

Recently, endmember extraction algorithms have exploited spatial input, which is a distinguishing characteristic of image data, as well as spectral information. The following includes recently proposed spectral and spatial based EEAs:

- Automated morphological endmember extraction (AMEE)

  The AMEE [53] algorithm exploits mathematical morphology to search the spatial neighborhood around each pixel for the most spectrally singular and highly mixed pixels. Each spectrally singular pixel is assigned an "eccentricity" value, which is calculated as the spectral angle distance between itself and the most highly mixed pixel in the same spatial neighborhood. The "eccentricity" values of the selected pixels are updated in an iterative process, where the size of the spatial neighborhoods increases to a pre-determined maximum value. The final endmember set is obtained by applying a threshold to the resulting gray-scale "eccentricity" image.

- Spatial-spectral endmember extraction (SSEE)

  The SSEE [54] is an extension of the PPI algorithm. A set of basis vectors that describes most of the spectral variability is determined by singular value decomposition (SVD) of subsets in an image. These vectors are used as "skewers" to generate a set of candidate endmember pixels. The spectrally similar endmember candidates, which are spatially related, are averaged, while the endmembers that are spectrally similar but spatially independent are retained.

- Spatial purity based endmember extraction (SPEE)

  In the SPEE [55] algorithm, spectral uniformity of pixels within a fixed kernel window around each pixel is examined to extract endmember candidates. Then, the graph-based spatial connectivity of these endmember candidates is utilized to vote for new spatially independent endmember candidates. Pixel and feature

level methods to determine the spectral purity are also compared in terms of statistical accuracy of spectral angle distance and processing time.

2.2 Dimensionality Reduction (DR) for Pixel Unmixing

Due to the high dimensionality and spectral redundancy of hyperspectral data, DR methods are widely used to decrease the computational cost as well as to improve the accuracy of the data analysis [56]. DR is usually achieved via either feature selection or feature extraction [57]. In classification, feature selection aims to identify a subset of the original features that maintains the useful information to separate the classes with highly correlated and redundant features. These methods are dependent on the properties of the input data as well as on classifier used, and require a criterion to judge the quality of each feature. Because integer programming based optimization methods are computationally expensive, heuristic methods that do not guarantee global optima are often unitized in practice [58]. Feature extraction methods project the original remotely sensed data to a new feature space, where a small set of features contains the vast majority of the original data's information [56, 58]. Although the resulting features are not directly related to the original spectral features, these methods are preferred for spectral unmixing because they are typically robust and computationally advantageous. The following provides a brief review of feature extraction methods that have widely been applied to hyperspectral remote sensing data.

2.2.1 Linear Feature Extraction Methods

In spectral unmixing, most endmember algorithms are implemented in either the original or a linearly transformed domain, which is appropriate when the sub-manifold is embedded linearly in the original space.

Principal Component Analysis (PCA) is widely used as a linear feature extraction methods in remote sensing [59]. The resulting linear combinations of the original

variables are mutually orthogonal features based on the second-order statistics of the samples. Although PCA has been successfully applied to many remotely sensed datasets, it is not the optimal feature extraction method since it does not always guarantee retention of the discriminatory information [60].

The maximum (or minimum) noise fraction (MNF) proposed by Green et al. [40] was first developed as an alternative to the PCA to address the issues of the data variance are not always related to image quality, and principal components from PCA are not necessarily ordered by image quality. The MNF transformation was derived as an analogue of the PCA and has all the properties of the PCA, but utilizes signal-to-noise ratio (SNR) to measure image quality, resulting in the equivalent of a noise-whitened PCA [61, 62].

2.2.2 Manifold Learning for Analysis of Hyperspectral Data

Neither PCA nor MNF are able to handle nonlinear characteristics often exhibited in hyperspectral data that cannot be characterized by second-order statistics [63]. To explore this issue, various manifold learning methods have been proposed by the machine learning community [64–68]. Common methods include isometric feature mapping (ISOMAP) [64], locally linear embedding (LLE) [65], kernel PCA (KPCA) [67], and local tangent space alignment (LTSA) [68]. They are generally characterized as 1) global methods (e.g., ISOMAP and KPCA) and 2) local methods (e.g., LLE and LTSA). Whereas the global manifolds preserve geometry at all scales and provide a faithful representation of the data's global structure [64], the local approaches have better representation of locally complex geometry and accommodation of submanifolds [65].

Although nonlinear manifolds based on ISOMAP, one of the most popular global methods, have been investigated for classification [8, 32–36, 61, 69] and unmixing [37–39], they are computationally intensive because of two bottlenecks associated with the shortest path algorithm for the geodesic distance calculation and multidimensional

scaling (MDS) eigenvalue computation [8, 61, 69–71]. To reduce the computational costs, a small number of points referred to as "*landmarks*" are often chosen to develop the manifold, and the remaining points are embedded in the space. Random selection is the most common for obtaining landmarks in practice. Random selection samples from the uniform distribution, providing an easily implemented representation of the global manifold geometry, it cannot reliably represent important local variability within the manifold. The "maxmin" method, which maximizes the minimum distance from the set of landmarks to all other non-landmark points, was proposed to reduce the uncertainty and improve the landmarks for classification [70]. In [71], a new landmark selection method motivated by the Least Absolute value Subset Selection Operator (LASSO) [72] was proposed. The solution of a regression problem minimized a regularized cost function. The selected landmark set found geometrically meaningful landmarks and preserved the overall manifold geometry.

In the hyperspectral remote sensing community, Bachmann et al. [8] addressed the high computational load of ISOMAP via the "*backbone*" approach. In this method, called EHN-ISOMAP, large numbers of samples are inserted into a representative backbone manifold obtained by a subset of data (randomly selected landmarks). While the method effectively reduced an order of computation, it may not fully span the original space if a landmark set is too small. In [69], intelligent landmark points on a global manifold were selected using the "minimum spanning tree cut" method, which tends to choose points from cluster boundaries, which benefited classification. Lower SStress values were obtained than by random selection and k-means cluster centers, and the classification accuracy approached that of non-landmark ISOMAP. Additional details of backbone methods are contained in Chapter 3, where the proposed spatial-spectral method is described.

2.3 Previous Research for Nonlinear Spectral Unmixing

The following is an overview of various nonlinear spectral unmixing studies in the last decade. Most of studies have concentrated on direct nonlinear unmixing models. Recently, manifold learning based unmixing studies have been raised.

2.3.1 Direct Nonlinear Unmixing Model Appraoches

Liu and Wu [28] compared diverse families of pattern recognition based data driven nonlinear unmixing methods such as multilayer perception (MLP), neural networks, adaptive resonance theory of pattern recognition, and regression trees with binary recursive partitioning. The comparison was conducted based on model parameters, computational complexity, prediction accuracy, and interpolation ability. ARTMMAP, which is the extension version of popular neural network model ARTMAP significantly outperformed ARTMAP and is also easy to implement. The MLP model resulted in stable outputs, but has a heavy computational overhead and many experimental trials.

In [29], a kernel-based method was demonstrated for unmixing. Results obtained using various kernels (e.g., linear, RBF, polynomial) and physics-based approaches were compared. The polynomial kernel resulted in the lowest reconstruction RMSE, but because of its approximation of the physics, it ultimately failed when the mixtures were highly nonlinear.

In [30], selection of the most highly informative samples for effectively training the neural architecture was discussed. This issue was also addressed by several algorithms for intelligent selection of training samples such as border-training, mixed-signature, and morphological-erosion algorithms. These algorithms were used to train a MLP which is a representative neural architecture for nonlinear spectral mixture analysis.

Altmann et al. [31] proposed a model which assumes that the pixel reflectances are nonlinear functions of pure spectral components contaminated by an additive white Gaussian noise. A polynomial post-nonlinear mixing model which is a flexi-

ble generalization of the standard linear mixing model was used to estimate these nonlinear functions. Optimal parameters associated with the model were estimated by a Bayesian approach. The model accurately modeled nonlinearities in mutiple synthetic and real hyperspectral images.

2.3.2 Manifold Learning Based Nonlinear Unmixing Approaches

Heylen et al. [37] employed manifold learning to exploit nonlinear mixing assumptions. The proposed algorithm is based on simplex volume maximization and uses the geodesic distances, computing the shortest path distance in the nearest neighbor graph along with global data manifold (e.g. the ISOMAP (Isometric feature mapping) approach). Although the differences between the linear and nonlinear models were difficult to quantify due to the limited ground reference data, the method captured the subtle differences of the absorption features better than linear approaches.

Geodesic distance is commonly used in manifold learning methods. However, geodesic distances commonly obtained from shortest path algorithms, may deviate significantly from the true shortest path in real datasets. In [38], Heylen and Scheunders revised geodesic distance calculation in the ISOMAP using a generalized bilinear model instead of the typical graph-based model. Multidimensional scaling associated with the ISOMAP method was recalculated using the true geodesic distances. The absolute difference by the proposed method outperformed the classical ISOMAP as well as the linear DR methods.

Hoang et al. [39] compared global and local manifold learning approaches to deal with nonlinearities in the DR step and integrated them into the several endmember extraction algorithms. The global manifold ISOMAP generally resulted in better statistical performance than the local approach LLE (Locally linear embedding), although with high computational overhead.

3. LANDMARK BASED MANIFOLD LEARNING FOR NONLINEAR UNMIXING

Nonlinear manifold learning-based spectral unmixing provides an alternative to direct nonlinear unmixing methods for representing nonlinearities inherent in hyperspectral data. Although manifolds can effectively capture nonlinear features in the DR stage, the computational overhead is excessive for large-scale remotely sensed data. Manifold approximation using a set of distinguishing points is commonly utilized to mitigate the computational burden, but selection of these landmark points is important for adequately representing the topology of the manifold.

In this chapter, a new robust landmark selection method for the pixel unmixing that exploits the spectral and spatial homogeneity in a local window kernel is proposed to address the highly variable results of randomly selected landmarks. This approach is incorporated in development of a hybrid manifold which exploits both characteristics of global and local manifold learning[1].

3.1 Landmark Selection for Unmixing of Hyperspectral Data

3.1.1 Global Manifold: Isometric feature mapping (ISOMAP)

ISOMAP has been successfully demonstrated for data representation and classification of hyperspectral data [8, 61, 69]. The ISOMAP method assumes that the

[1]This chapter is extracted from [73] J. Chi and M. M. Crawford. Landmark selection using homogeneity on nonlinear manifolds for unmixing hyperspectral data. *Proc. 2012 IEEE Intl Geoscience and Remote Sensing Symposium (IGARSS)*, Munich, Germany, July 22-27, pp. 1373-1376, 2012 and [74] J. Chi and M. M. Crawford. Selection of landmark points on nonlinear manifolds for spectral unmixing using local homogeneity. *IEEE Geoscience and Remote Sensing Letters*, 10(4): 711-715, 2013.

nearest neighbors linearly form the local feature space, and then connecting these piecewise linear spaces via geodesic distances can globally embed the data into non-linear transformed coordinates [34, 64]. Given data samples in a hyperspectral data matrix $[\mathbf{X} : \mathbf{x}_i, i = 1, \ldots, N] \in \mathbf{R}^D$, where D is the original dimension and N is the number of samples, the goal of DR is to embed high dimensional samples in a lower d dimensional space $\mathbf{R}^d$, where typically, $d << D$, while preserving the geometry as faithfully as possible [64,70]. ISOMAP, which attempts to recover the original hidden space of the data, consists of three steps:

1. Constructing the nearest neighborhood graph $\mathbf{G}$ of $\mathbf{X}$ using the k-nearest neighbor rule, where each edge is weighted by the Euclidean distance between all pairs of data points, i.e., $O(DN^2)$.

2. Computing the shortest path matrix in $\mathbf{G}$ to approximate the geodesic distance using either Floyd's algorithm $O(N^3)$ or Dijkstra's algorithm $O(kN^2\log N)$.

3. Solving the multidimensional scaling (MDS) eigenvalue problem to find a new embedding space, i.e., $O(N^3)$.

3.1.2 Landmark ISOMAP (L-ISOMAP)

The high computational overhead associated with the shortest path matrix and MDS eigenvalue computation is especially problematic for large-scale remote sensing datasets [8, 61, 64, 69, 70]. The computational complexity in general depends not only on the dimensionality but also on the number of data points. The computational cost of the ISOMAP is quadratic in the number of points [71]. Use of a small set of points, referred to as "*landmarks*," was the motivation for reducing the computational complexity of ISOMAP [70, 71]. Using a set of n landmarks, denoted by $[\mathbf{L} : \mathbf{l}_i, i = 1, \ldots n] \in \mathbf{R}^D$, the distances between the N data points and n landmark points require an $n \times N$ sub-matrix instead of the $N \times N$ distance matrix of ISOMAP, where typically $n << N$. The L-ISOMAP method can be described in three steps:

1. Selecting a set of n landmarks.

2. Applying classical MDS to find an embedding of n landmark points using the $n \times n$ matrix of distances between each landmark point.

3. Finding a d-dimensional embedding of the N data points using the $n \times N$ of distances between the landmark points and the remaining points.

As a consequence, the order of the shortest path and the MDS computations can be reduced from $O(kN^2 \log N)$ and $O(N^3)$ to $O(knN \log N)$ and $O(n^2 N)$, respectively [8, 69, 70]. Although a minimum of $d+1$ landmarks is required, it is generally advisable to select more than the strict minimum to maintain stability and to increase representation.

3.1.3 Landmark Selection Methods

Many techniques have been proposed to choose a set of landmarks. Random selection, which often results in uniform approximation of the given data distribution, is the easiest practical way to select a landmark set and is generally advisable for smoothly varying manifolds; however, it induces uncertainty into the resulting manifold and is not controllable. So, it is unlikely to be the best strategy, especially in many practical applications with complex geometry. The constructed set may contain many redundant samples, motivating researchers to develop better sampling strategies [69, 70].

A. Maxmin

The maxmin method is designed to rapidly seek spectrally extreme values in the dataset and choose the landmark points according to the following steps:

1. Randomly choose $1 \leq l \leq n$ initial seed points, adding them to the set of landmarks $\mathbf{L}$ and removing them from the set of data points $\mathbf{X}$.

2. For $\mathbf{x}_i \in \mathbf{X}$ and $\mathbf{l}_j \in \mathbf{L}$, let

$$\mathbf{D}_i = \min_{j=1:\|\mathbf{L}\|} \{\text{dist}(\mathbf{x}_i, \mathbf{l}_j)\} \tag{3.1}$$

3. Let $\mathbf{D}_k$ be the maximum of $\{\mathbf{D}_i\}$. Add $\mathbf{x}_k$ to $\mathbf{L}$ and remove it from $\mathbf{X}$.

4. Repeat Step 2 and Step 3 until the number of desired landmarks is found, $\|\mathbf{L}\| = n$.

This method has additional operations of $O(nN)$ compared with random selection, but it has the advantage of creating a much smaller landmark set than is needed by random selection to obtain approximately the same results [70]. The maxmin method also has a corresponding danger of selecting outliers especially in hyperspectral remotely sensed data.

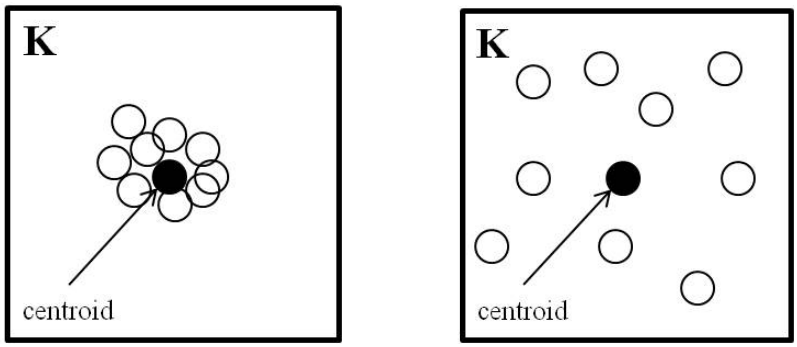
B. Local Homogeneity (LH)

Plaza et al. [53] exploited mathematical morphology to discriminate the most spectrally and spatially singular pixels, as well as the mixed pixels in a window kernel. However, the dilation and erosion operations require an additional $O(p^3)$ computation for each pixel, where p is the kernel size. This is counter to the goal of the use of landmarks in ISOMAP.

If the p^2 pixels in a kernel are highly correlated, they are spectrally singular (e.g., spectrally pure or the same mixture) at least within a local area (see Fig. 3.1(a)). Otherwise, the pixels within the spatial kernel might be highly mixed (see Fig. 3.1(b)). The spectral consistency in spatial kernel K can be written as follows:

$$\text{CORR}(\mathbf{x}_1, \mathbf{x}_2) = \frac{\langle \mathbf{x}_1, \mathbf{x}_2 \rangle}{\|\mathbf{x}_1\| \, \|\mathbf{x}_2\|} \tag{3.2}$$

$$\text{Spectral Consistency} = \frac{\sum_{(s,t)\in K} \text{CORR}(\mathbf{x}_{s,t}, \widehat{\mathbf{x}})}{p^2} \tag{3.3}$$

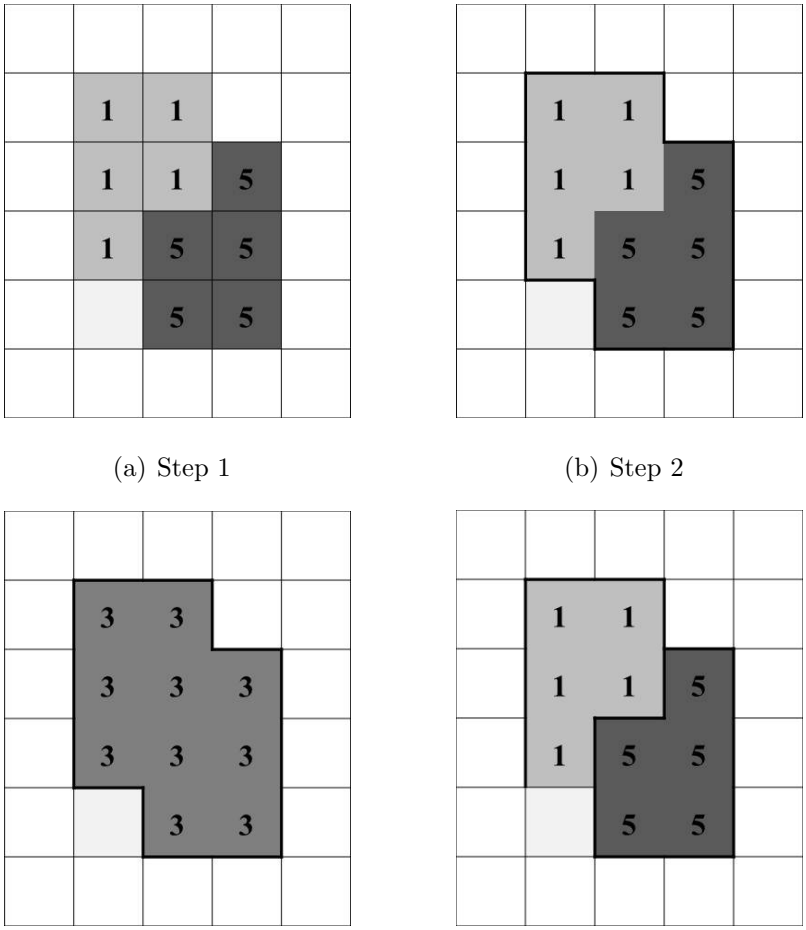


(a) Homogeneous (b) Mixed

Figure 3.1. Determination of local homogeneity: (a) spectrally singular pixels and (b) spectrally mixed pixels in K

where $\mathbf{x}_{s,t}$ is a pixel vector with spatial coordinates (s,t) in K, and the mean spectral vector value of K is denoted by $\widehat{\mathbf{x}}$. Otsu's threshold method, which performs automatic histogram shape-based thresholding, can be used to determine the homogeneity of pixels [75].

In remotely sensed data, pixels adjacent to homogeneous pixels often have similar spectral characteristics. Taking into account geometric connectivity inhibits the selection of spatially localized landmark points, thereby increasing the spatial diversity. As shown in Fig. 3.2(a), for example, homogeneous pixels were identified by local spectral consistency and have two spectral "values". Using all the homogeneous pixels as landmarks may result in loss of spectral information due to both the limited number of resulting landmarks and their spatial proximity. Increasing spectral and spatial diversity results in the selection of additional landmarks. However, if the geometric connectivity described in [55] is considered, those spectrally different homogeneous pixels may be considered to be a single large group due to the smoothing by the kernel and spatial proximity (see Fig. 3.2(b)). This may result in selection of one inappropriate value from a large homogeneous group since the candidate determination step in SPEE selects the mean signature from each group (see Fig. 3.2(c)). However, more than one subspace of pixels can exist in a large group. The proposed



(a) Step 1 (b) Step 2

(c) Step 3 (d) Step 4

Figure 3.2. Spatial proximity: (a) pixels identified as homogeneous by spectral proximity, (b) geometric connectivity in SPEE, (c) determination of representative value by SPEE (Spatial Purity based Endmember Extraction), and (d) selection of representative values using the proposed method

method chooses multiple values in the large patch based on spectral and spatial proximity since spectral correlation is re-considered prior to connecting the points in the graph (see Fig. 3.2(d)). In this study, the pixel that is the most spectrally similar to the mean spectral signature of the spectrally homogeneous patch is selected. Additionally, proper selection of the kernel size and shape can remove outliers and thus result in a smoother, more representative manifold.

3.2 Hybrid Manifold Learning to Exploit Global and Local Characteristics

3.2.1 Local manifold: Locally Linear Embedding (LLE)

LLE [65] is a widely used local manifold method which is based on simple geometric intuition. It characterizes the local geometry of a locally linear patch of the manifold by linear coefficients in a lower dimensional space that reconstructs each data point from its neighbors. While ISOMAP connects points by traversing a graph, LLE finds a set of weights by identifying the nearest neighbors based on a linear distance, and then performs local linear interpolations that closely approximate the data. Since the weights are invariant to rotation, rescaling and translation of the data points and their neighbors, LLE has better representation of complex geometry and of sub-manifolds which exist in the dataset [64,65,76]. LLE generally consists of the following steps [65, 76]:

1. For each data point $\mathbf{x}_i$ in $[\mathbf{X} : \mathbf{x}_i, i = 1, \ldots, N] \in \mathbf{R}^D$, find the k-nearest neighbors $\mathbf{x}_j$.

2. Find the weight matrix $\mathbf{W}$ which minimizes the reconstruction error for each point $\mathbf{x}_i$ from its neighbors,

$$\epsilon(\mathbf{W}) = \sum_{i=1}^{N} \left\| \mathbf{x}_i - \sum_{j \neq i} \mathbf{w}_{ij} \mathbf{x}_j \right\|^2 \tag{3.4}$$

   where $\mathbf{w}_{ij}$ denotes the reconstruction weight of $\mathbf{x}_i$ from its j-th neighbor $\mathbf{x}_j$ and is constrained such that $\sum_j \mathbf{w}_{ij} = 1$.

3. Compute the embedded coordinates $[\mathbf{Y} : \mathbf{y}_i, i = 1, \ldots N]$ which minimize the reconstruction error using the coefficients from Step 2,

$$\Phi(\mathbf{Y}) = \sum_{i=1}^{N} \left\| \mathbf{y}_i - \sum_{j \neq i} \mathbf{w}_{ij} \mathbf{y}_j \right\|^2 \tag{3.5}$$

   subject to the constraints that $\sum_i \mathbf{y}_{ij} = 0$ for each j and $\frac{1}{N} \sum_i \mathbf{y}_i \mathbf{y}_j^T = \mathbf{I}$.

3.2.2 Hybrid Manifolds (HMs) for Hyperspectral Data

In [8], a reconstruction approach that utilized a backbone developed using landmarks was proposed to further reduce the computational overhead for developing a manifold. The method was based on ISOMAP to develop the backbone as well as a large landmark set which was approximately 10-33% of the original data points. The approximation to insert the points not used in the backbone into the backbone manifold is similar to the reconstruction principle used in LLE.

Although the LLE itself is computationally more efficient than ISOMAP, only the k-nearest neighbors contribute to the reconstruction, which implies that the LLE preserves the local structure of the geometry, but not the global structure. In addition, anomalies which are potentially problematic for remote sensing data may affect the manifold recovery. The proposed "*hybrid manifold* (HM)" preserves both the global and local structures in the data using globally selected backbone points and linearly embedding data points based on the backbone. The backbone includes spectrally and spatially representative pixels that have been identified by the proposed LH method. Our HM has three key characteristics which address the weaknesses of the global and local manifolds:

- Applying the LH method, well-separated global landmarks are used as backbone points to develop the backbone manifold and contribute to preservation of global structure. Further, they preserve a meaningful local structure while including complex spectral geometry, and they contribute to the reconstruction of the remaining points.

- The HM method is robust to anomalies because the backbone points chosen by the LH method neglect outliers. This reduces negative effects on manifold recovery which are problematic for local manifolds.

- The HM is computationally efficient compared to LLE as well as ISOMAP.

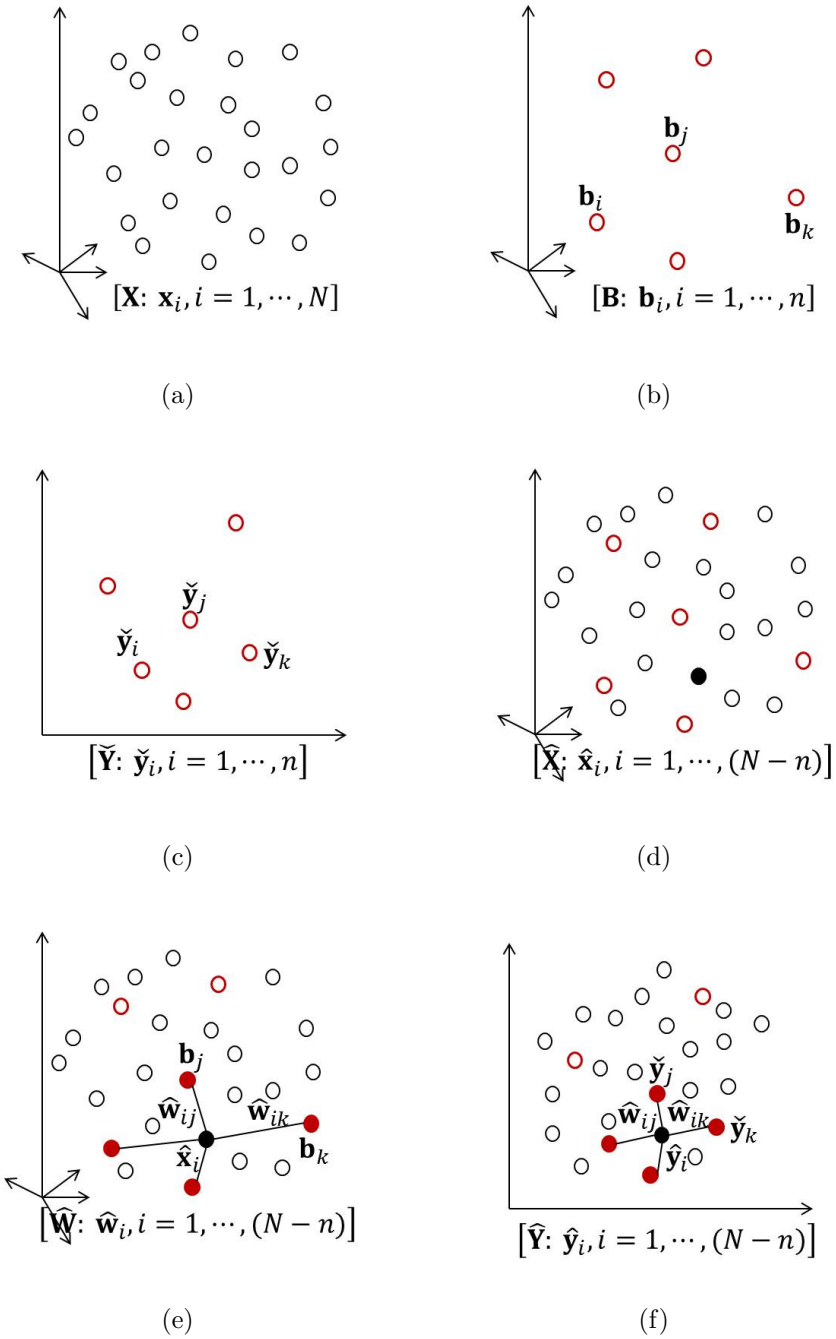


Figure 3.3. Steps of hybrid manifold: (a) original data points, (b) selected backbone points, (c) updated backbone via LLE, (d) remaining points, (e) reconstruction weights, and (f) embedding phase

The key idea of the proposed method is to embed points on a backbone developed from landmarks that retain information on the local structure. This indicates that selection of relevant backbone points is a key component of this algorithm. The HM can be implemented with the following steps:

1. Select a set of backbone points $[\mathbf{B} : \mathbf{b}_i, i = 1, \ldots, n]$ by the proposed LH landmark selection method described in the previous section (Fig. 3.3(b)).

2. Perform classical LLE on the selected points in Step 1 to build a backbone in a lower dimensional embedding space, $\left[\check{\mathbf{Y}} : \check{\mathbf{y}}_i, i = 1, \ldots, n\right]$ (Fig. 3.3(c)).

3. For the remaining points $\left[\hat{\mathbf{X}} = \mathbf{X} \backslash \mathbf{B} : \hat{\mathbf{x}}_\mathbf{i}, i = 1, \ldots, (N-n)\right]$, compute distances from $\hat{\mathbf{X}}$ to $\mathbf{B}$, and then find the k-nearest backbone points.

4. Obtain reconstruction weights, $\left[\hat{\mathbf{W}} : \hat{\mathbf{w}}_i, i = 1, \ldots, (N-n)\right]$ resulting from $\hat{\mathbf{x}}_i$ to its neighbors, $\mathbf{b}_j$ by minimizing the cost function (Fig. 3.3(e)):

$$\epsilon(\hat{\mathbf{W}}) = \sum_{i=1}^{N-n} \left\| \hat{\mathbf{x}}_i - \sum_{j \neq i}^{k} \hat{\mathbf{w}}_{ij} \mathbf{b}_j \right\|^2 \tag{3.6}$$

   subject to $\sum_j \hat{\mathbf{w}}_{ij} = 1$.

5. Find d dimensional embedding coordinates, $\left[\hat{\mathbf{Y}} : \hat{\mathbf{y}}_i, i = 1, \ldots, (N-n)\right]$ for the remaining points $\hat{\mathbf{X}}$ using the reconstruction weights $\hat{\mathbf{W}}$ (Fig. 3.3(f)):

$$\hat{\mathbf{y}}_i = \hat{\mathbf{w}}_1 \check{\mathbf{y}}_1 + \hat{\mathbf{w}}_2 \check{\mathbf{y}}_2 + \cdots + \hat{\mathbf{w}}_k \check{\mathbf{y}}_k \tag{3.7}$$

6. Combine $\check{\mathbf{Y}}$ and $\hat{\mathbf{Y}}$.

LLE is computationally efficient in itself compared to ISOMAP. Table 3.1 lists the computational complexity metrics for comparison of LLE and the HM. There is no significant improvement in the stage of reconstruction weight computation, but the k-nearest neighbor selection and the embedding phase which are the most computationally intensive parts of the LLE require less computational overhead than classical LLE because of $n << N$.

Table 3.1 Computational Complexity Comparison

Function	LLE	Hybrid Manifold	
		Backbone Development	Embedding points
k-nn selection	$O(DN^2)$	$O(Dn^2)$	$O(DNn)$
Reconstruction weights	$O(Dk^3N)$	$O(Dk^3n)$	$O(Dk^3(N-n))$
Embedding	$O(dN^2)$	$O(dn^2)$	$O(d(N-n))$

3.3 Experimental Results

3.3.1 Data Description: AVIRIS Cuprite

The NASA/JPL AVIRIS (Airborne Visual and Infra-Red Imaging Spectrometer) instrument collected hyperspectral image data over Cuprite, Nevada from 1990 to 1995 for mapping minerals. Many studies related to spectroscopy have been conducted using both image and field data from this testbed. The AVIRIS Cuprite dataset is one of the most popular datasets in the spectral mixture analysis group since it helps in an understanding of the environment and the complexity of the surficial mineralogy [77]. A 300×250 subset with $17m$ spatial resolution, and spectral range from approximately $1990 - 2480nm$, has been selected to evaluate the proposed methods in this study (Fig. 3.4). Each mineral has a specific crystal structure as well as subtle changes in this wavelength range [78].

3.3.2 Endmember Identification via N-FINDR

The dataset potentially contains 9 endmembers estimated by the HySime virtual dimensionality algorithm, which is based on signal subspace identification by minimum error [79]. The dataset has been shown to contain at least 5 minerals, which are alunite, buddingtonite, calcite, kaolinite, and muscovite, and they are used for eval-

uations. Their spectral reflectance and absorption features derived from the USGS spectral library are shown in Fig. 3.5.

In this study, N-FINDR, which finds a simplex spanned by q+1 endmembers embedded in a lower data space with a maximal volume [42], is used to find endmembers. It first maps the original data points into a lower q dimensional space using a dimensional reduction method and then the simplex volume formed by randomly selected seed points is computed based on the matrix:

$$\mathbf{M} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{e}_0 & \mathbf{e}_1 & \cdots & \mathbf{e}_q \end{bmatrix} \tag{3.8}$$

The volume of the resulting simplex spanned by the endmembers $[\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_q]$ is proportional to the determinant of $\mathbf{M}$:

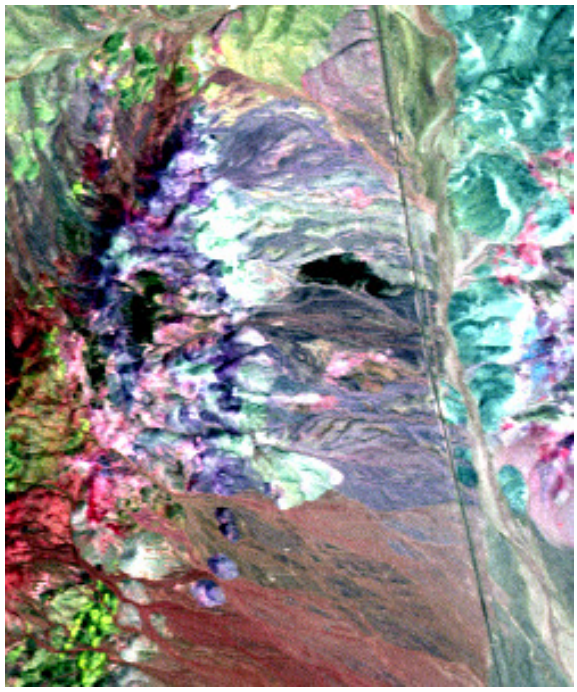


Figure 3.4. AVIRIS Cuprite dataset

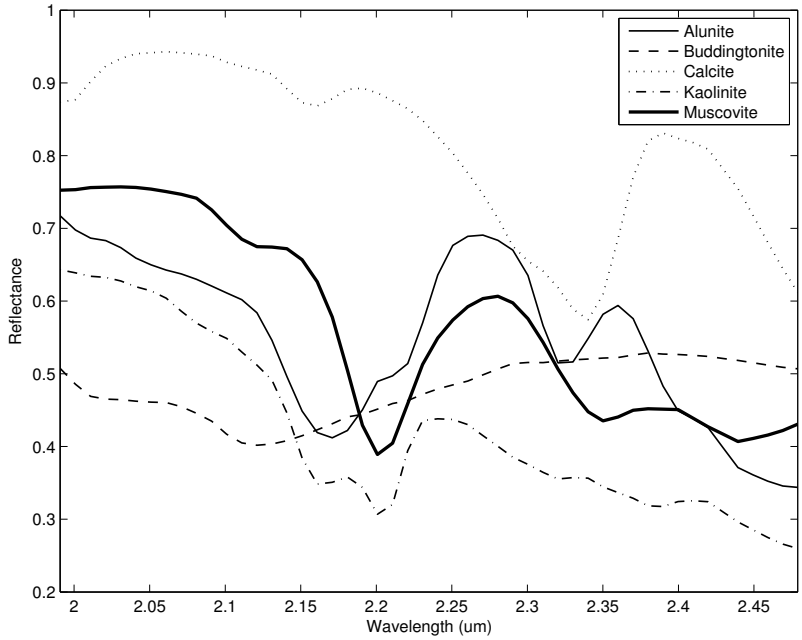


Figure 3.5. Spectral reflectance of ground reference from USGS spectral library

$$V(\mathbf{M}) = \frac{1}{q!}\,\mathrm{abs}(\det(\mathbf{M})) \tag{3.9}$$

To update the initial estimate of the volume, the volume is computed for every pixel in each endmember position by replacing the pixel when a larger simplex is found, and then recalculating the volume. This procedure is continued until no larger simplex is found, as follows:

$$\arg \max_{(\mathbf{e}_0,\mathbf{e}_1,\ldots,\mathbf{e}_q)} V(\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_q) \tag{3.10}$$

N-FINDR requires dimensionality reduction since $\mathbf{M}$ must be a square matrix for its determinant to exist. In this study, the endmembers are extracted from PCA as a linear approach and extracted along the nonlinear manifolds ISOMAP, L-ISOMAP, LLE, and HM as manifold learning based approaches.

3.3.3 Experimental Strategy

A set of 324 landmarks was selected using the landmark selection method described in the previous sections and used in L-ISOMAP and for backbone points in the HM. The number of resulting landmarks obtained by the LH method was determined based on spectral and spatial proximity, and the same number of landmarks was selected using random selection and the maxmin approach. For the proposed method, a 3×3 window kernel was adopted to evaluate local spectral consistency. N-FINDR arbitrarily chose the initial seed points, which were fixed for all the experiments to minimize variations in the results. Once the endmembers were chosen, fractional abundances were estimated as linear combinations of extracted endmembers for each pixel using a fully constrained least square unmixing algorithm as described in Section 1.4.1

3.3.4 Distribution of Landmark Points

Fig. 3.6 illustrates the distributions of selected landmark (backbone) points of the dataset on L-ISOMAP and HM spaces identified by the random selection, the maxmin approach, and the proposed LH method. Randomly selected landmarks are neither spectrally nor spatially meaningful, even if they are visually well distributed on the scatter plots (see Fig. 3.6(a), 3.6(d)). The landmark points identified by the maxmin method are well distributed over the scatter plots (see Fig. 3.6(b), 3.6(e)). Since the method maximizes spectral distances between landmark points, extracted points should be spectrally distinguishable and include extreme points. However, outliers or small isolated areas can also be included in the landmarks. As shown in the scatter plot, many points are actually selected from outside of the manifold space defined by the samples. In the proposed method, selected points are well distributed in the transformed space (see Fig. 3.6(c), 3.6(f)). These pixels are both spectrally and spatially homogenous and have meaningful values which are representative pixels of the image. As shown in the scatter plot, these landmark points are mainly distributed

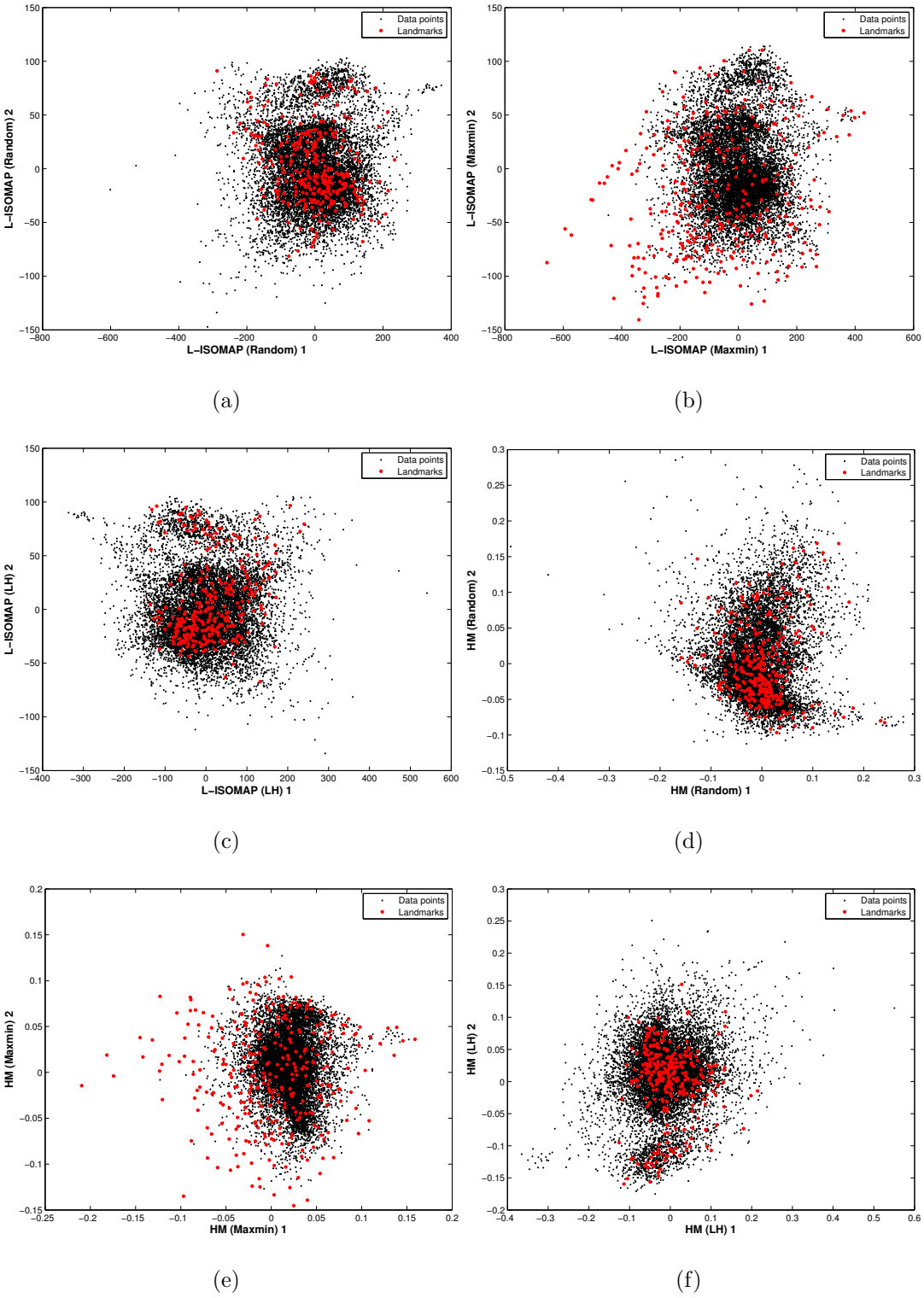


Figure 3.6. Distribution of extracted landmarks using: (a) random selection, (b) maxmin method, (c) LH method on L-ISOMAP, and (d) random selection, (e) maxmin method, (f) LH method on HM

within clusters of points in the manifold representing local geometry, indicating that the method is not sensitive to outliers.

As shown in Fig. 3.6(a)-3.6(c), since L-ISOMAP recovers ISOMAP's space with good fidelity, the resulting scatter plots are similar, although they have the opposite sign. However, point clouds of the HM (Fig. 3.6(d)-3.6(f))are totally different as they depend on the set of backbone points used to develop the manifold. Clearly, selection of backbone points is a critical component of the HM.

3.3.5 Evaluation of Endmembers via Spectral Angle Distance

Ten replications of experiments for each DR method were conducted, and the individual, mean and standard deviations of 5 well-known minerals are listed in Table 3.2. Based on the assumption that the correct endmember has the minimum SAD with ground reference, as shown in Table 3.3, PCA had a significant problem with kaolinite and muscovite, which are classified into the same image extracted endmember since they have similar absorption features near 2200nm (see Fig. 3.5). However, all the manifold based methods identified all the endmembers, and had lower average SAD.

Since a large number of points are approximated using the relatively small number of landmarks, the manifold approximations may be significantly affected by the landmarks and lead to different unmixing results. As shown in Table 3.2, random selection has the highest values of mean and standard deviations of SAD. It should be noted that while these differences seem to be relatively small in value, there were significant differences in fractional abundance maps. Fig. 3.7 illustrates variations of the random selection in the fractional abundances of kaolinite resulting from multiple experiments.

The manifold developed by the maxmin method, which is sensitive to data associated with extremes, can result in selecting outliers as endmembers. More useful endmembers may also be missed due to the limited number of endmembers. The proposed LH method smoothes pixel values using the window kernel, thereby mitigating

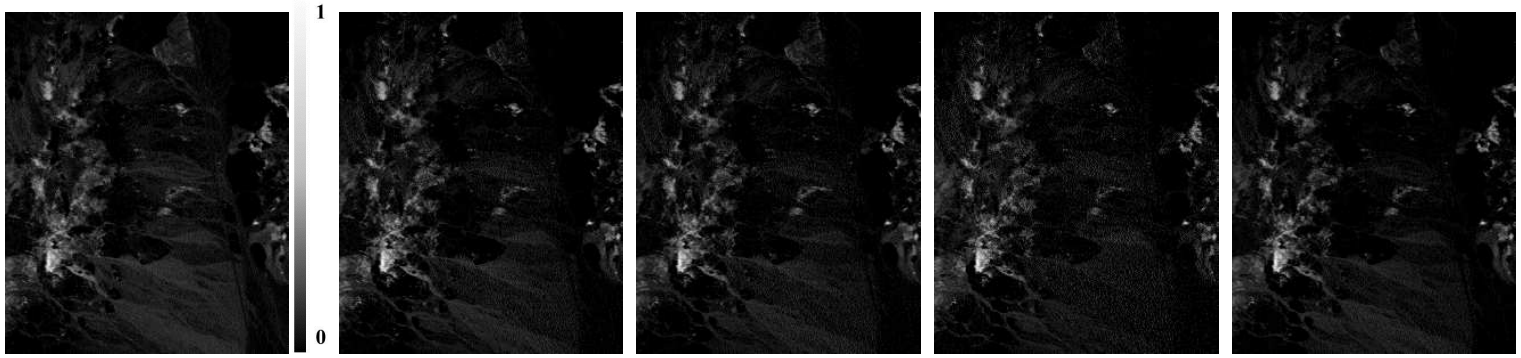
the impact of anomalies or small objects which would not contain mixed signatures anyway. The L-ISOMAP and HM by the LH method have smaller errors than those of the other approaches since the landmarks identified by the LH method played a role in developing more representative manifolds than other methods. Further, while the maxmin approach introduces variability associated with arbitrary selection of the initial seeds, the LH method does not introduce randomness.

Table 3.2 Comparison of spectral angle distance

Spectral Angle	PCA	ISOMAP	L-ISOMAP			LLE	HM (LH)
			Random	MaxMin	LH		
Alunite	**0.0546**	0.0741	0.0630 (0.0051)	0.0700 (0.0091)	0.0614	0.0597	0.0678
Buddingtonite	0.1389	0.1602	0.1356 (0.0055)	**0.1288** (0.0035)	0.1387	0.1466	0.1434
Calcite	**0.0668**	0.0927	0.0822 (0.0000)	0.0822 (0.0000)	0.0822	0.1019	0.1023
Kaolinite	0.1184	**0.0598**	0.0948 (0.0375)	0.0900 (0.0316)	**0.0598**	**0.0598**	0.0613
Muscovite	0.1081	**0.0796**	0.1064 (0.0078)	0.1008 (0.0068)	0.1127	0.0863	**0.0796**
Average	0.0974	0.0933	0.0964	0.0944	**0.0910**	**0.0908**	**0.0909**

Table 3.3 Spectral angle differences between extracted endmembers using PCA based method and ground reference

	E1	E2	E3	E4	E5	E6	E7	E8	E9
Alunite	0.1537	0.1741	0.1436	**0.0546**	0.1659	0.1927	0.1249	0.1241	0.1717
Buddingtonite	0.1751	0.2586	**0.1389**	0.2498	0.1606	0.2317	0.2035	0.2464	0.2927
Calcite	0.1064	0.1467	0.1618	0.2128	0.1085	**0.0668**	0.1373	0.1435	0.1582
Kaolinite	0.2152	0.2093	0.2536	0.1704	0.2351	0.2103	0.1765	0.1356	**0.1184**
Muscovite	0.1647	0.1662	0.2168	0.1780	0.1901	0.1762	0.1473	0.1081	**0.0797**



(a) #1 (0.0613) (b) #2 (0.0598) (c) #3 (0.0613) (d) #4 (0.0598) (e) #5 (0.1313)

(f) #6 (0.0613) (g) #7 (0.1288) (h) #8 (0.1033) (i) #9 (0.1313) (j) #10 (0.1493)

Figure 3.7. Variations of abundance maps of kaolinite by L-ISOMAP (Random)

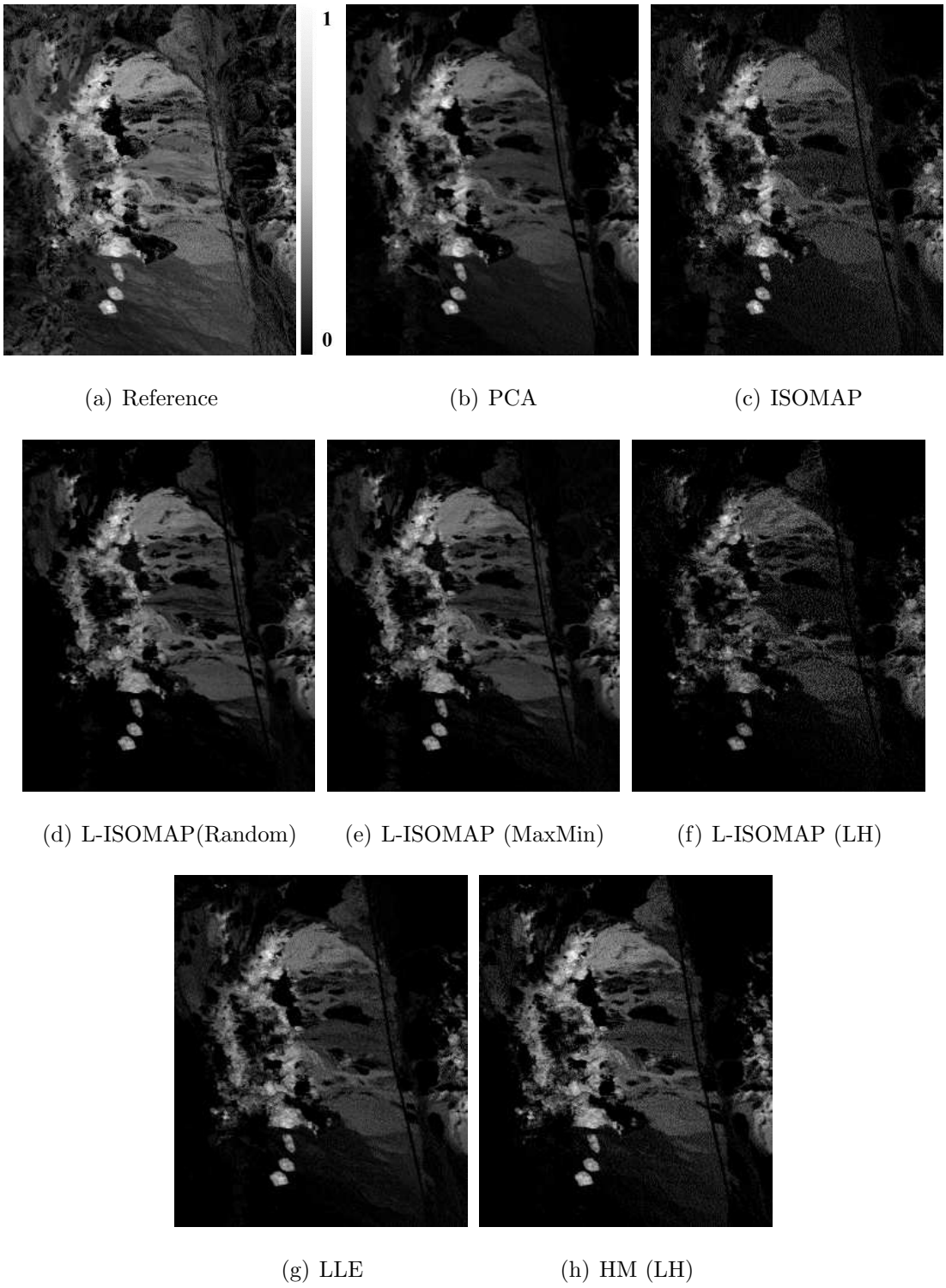
3.3.6 Evaluation of Endmembers via Fractional Abundance Map

Visual inspection of fractional abundance maps is useful for evaluating unmixing quality. For additional quantitative and qualitative analysis, reference fractional abundance maps were developed using USGS library spectra. However, these spectra are difficult to associate with the remote sensing data since they were acquired under different conditions from the image data. For these reasons, the pixels that have the smallest SAD with the library spectra were assumed to be the reference endmembers on the image scale, thereby fractional abundance maps based on the image derived reference were created using a fully constrained least square unmixing method.

Individual and average correlations with the reference are listed in Table 3.4, and Fig. 3.8-3.12 illustrate fractional abundance maps. Similar to SAD comparisons, manifold approaches are more highly correlated with the reference than the PCA-based method; in particular, much higher fidelity with the reference was observed in kaolinite and muscovite, which were problematic in the PCA-based method. Abundance maps obtained by ISOMAP generally exhibit better visual agreement with the reference and lower variability for all the endmembers. In terms of the SAD of the endmembers, LLE slightly outperformed ISOMAP, but not in the visual inspection. Since the remaining endmembers that were not used for the evaluations might be inaccurate, this resulted in poor visual agreement. Visual quality of the fractional abundances generally corresponded to the correlations with the reference. For all minerals, however, manifold approximation methods such as random selection (d), maxmin (e) and the LH method (f) of the L-ISOMAP, and the HM (LH) (h) method underestimated the fractional abundances compared to the classical ISOMAP (c) and LLE (g) in Fig. 3.8-3.12. The sensitivity of the maxmin method to outliers is shown in the abundance maps with very bright values over small areas especially in Fig. 3.10-3.12(e). The kernel filter in the proposed LH method mitigated the impact of these values; hence, the abundance maps (see Fig. 3.8-3.12) generally exhibited better visual agreement with the reference as well as higher correlation.

It is difficult to generalize the comparison of global and local manifolds since LLE yielded better results in the SAD comparison, but ISOMAP produced higher fidelity fractional abundances. Although the landmarks successfully solved the high computational overhead of ISOMAP, the results depend on which landmark points are selected. This is particularly critical for the HM since it computes the weights of the remaining points from the backbone points. The spectrally and geometrically well-separated points by the LH contribute to the preservation of the global and local geometry.

Therefore, L-ISOMAP and HM by the LH method had good results in both the endmembers and fractional abundance maps. This indicates that the proposed LH method was successfully incorporated into both the L-ISOMAP and HM methods. Additionally, although LLE itself is computationally efficient compared to ISOMAP, the computational overhead of the HM improved markedly.

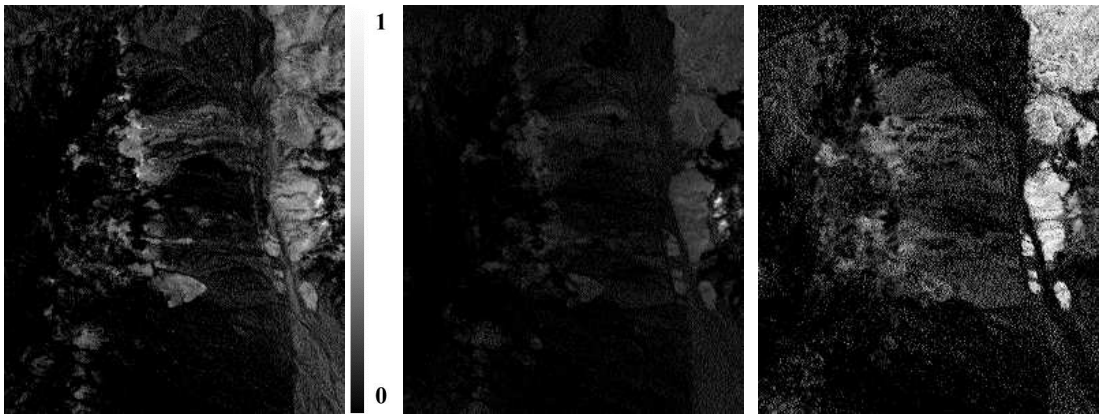


(a) Reference (b) PCA (c) ISOMAP

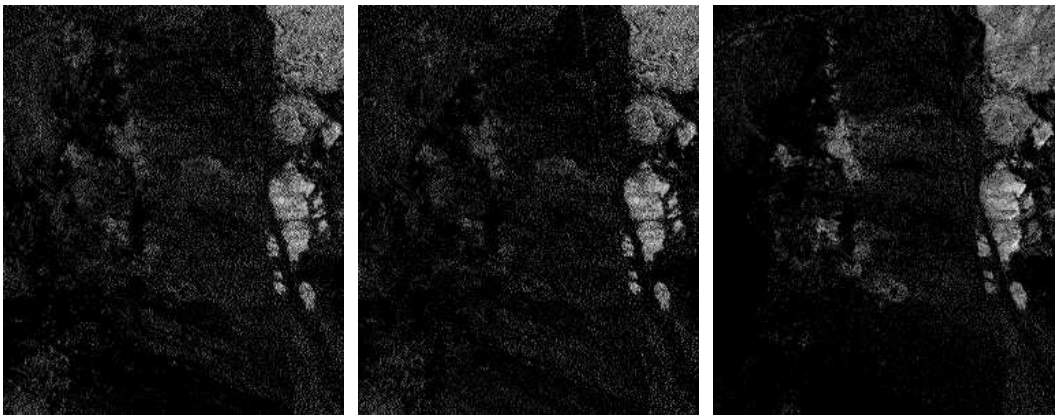
(d) L-ISOMAP(Random) (e) L-ISOMAP (MaxMin) (f) L-ISOMAP (LH)

(g) LLE (h) HM (LH)

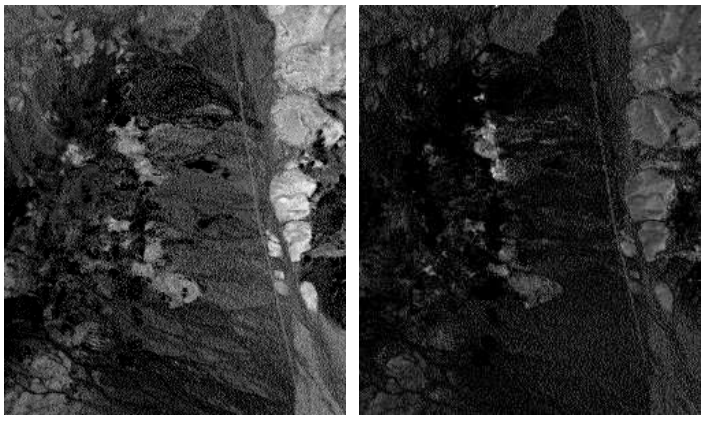
Figure 3.8. Abundance maps of Alunite



(a) Reference (b) PCA (c) ISOMAP

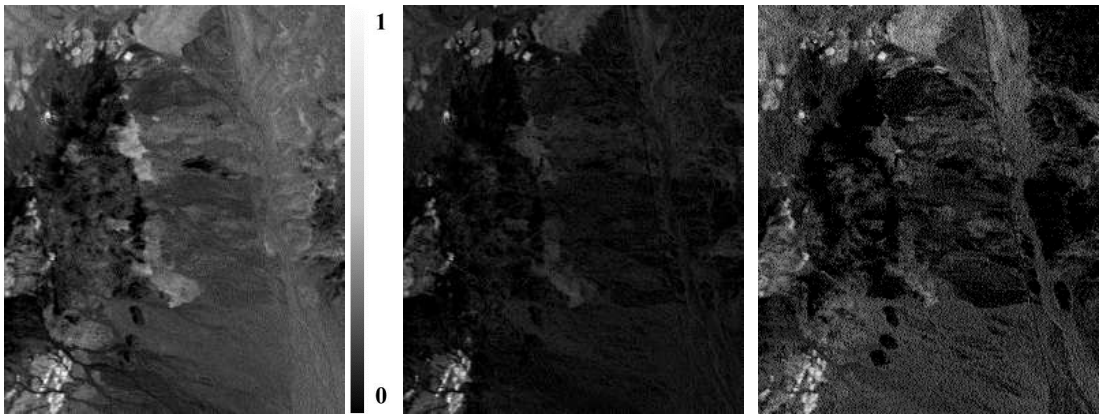


(d) L-ISOMAP(Random) (e) L-ISOMAP (MaxMin) (f) L-ISOMAP (LH)

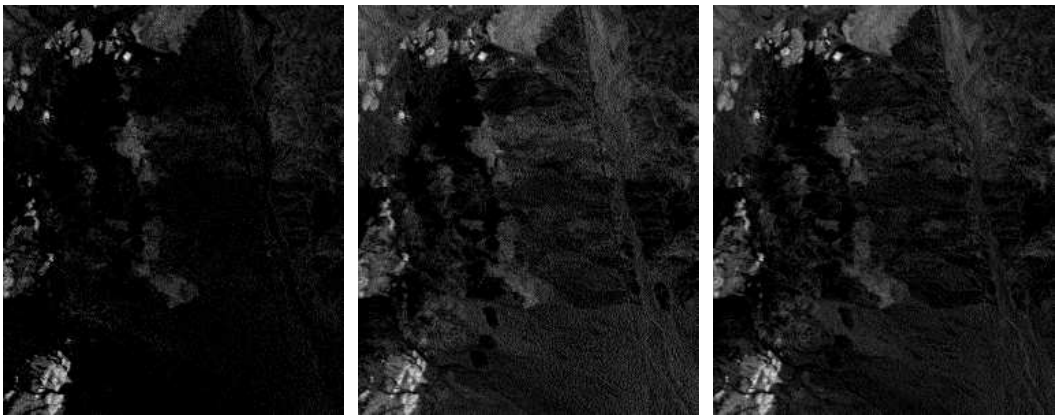


(g) LLE (h) HM (LH)

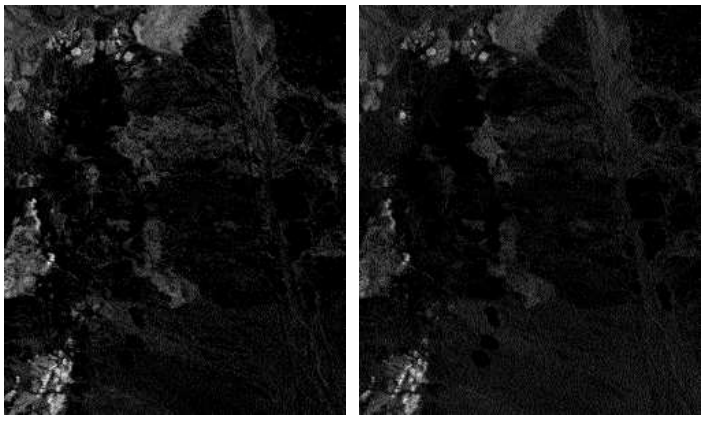
Figure 3.9. Abundance maps of Buddingtonite



(a) Reference (b) PCA (c) ISOMAP

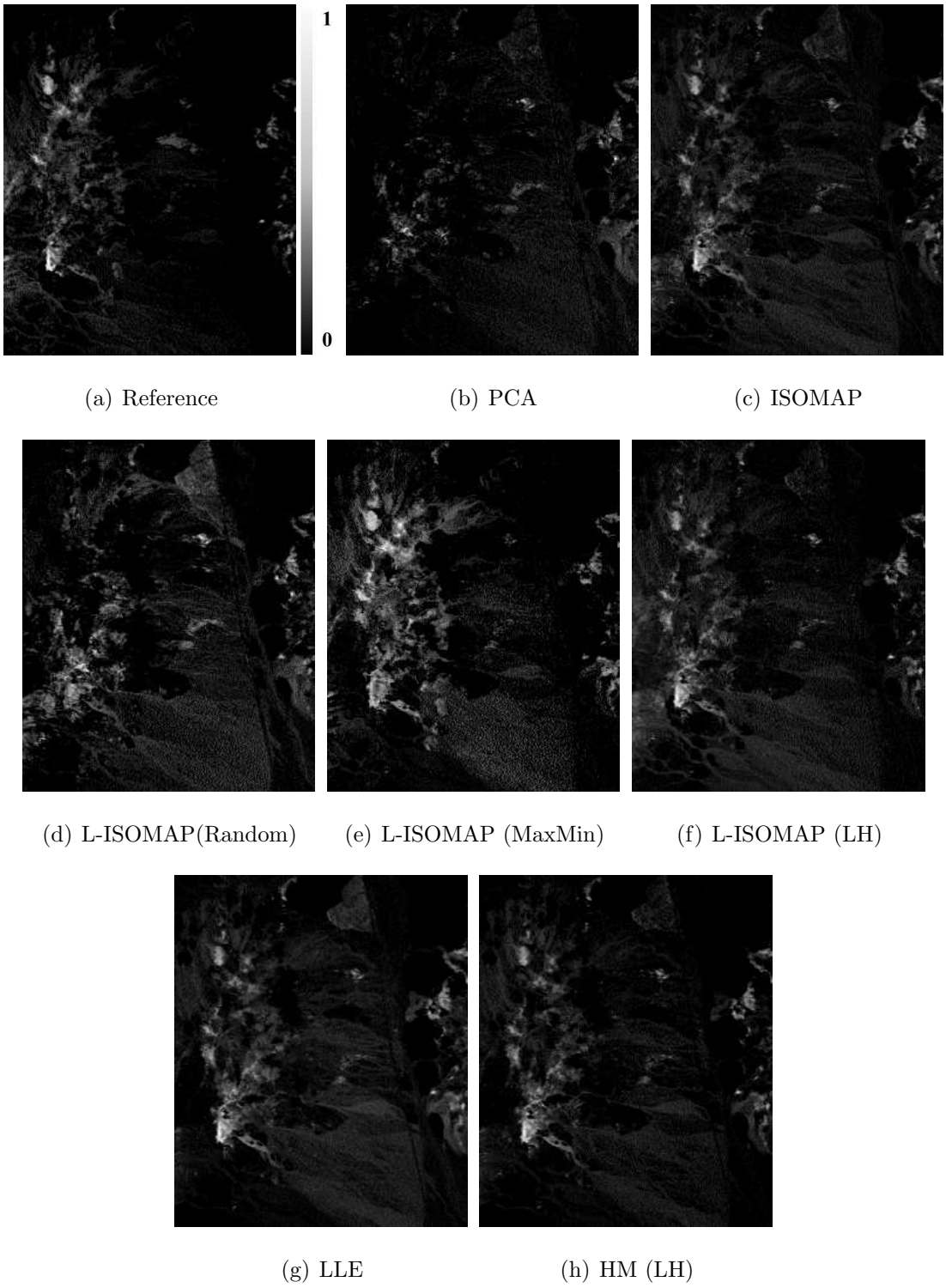


(d) L-ISOMAP(Random) (e) L-ISOMAP (MaxMin) (f) L-ISOMAP (LH)



(g) LLE (h) HM (LH)

Figure 3.10. Abundance maps of Calcite

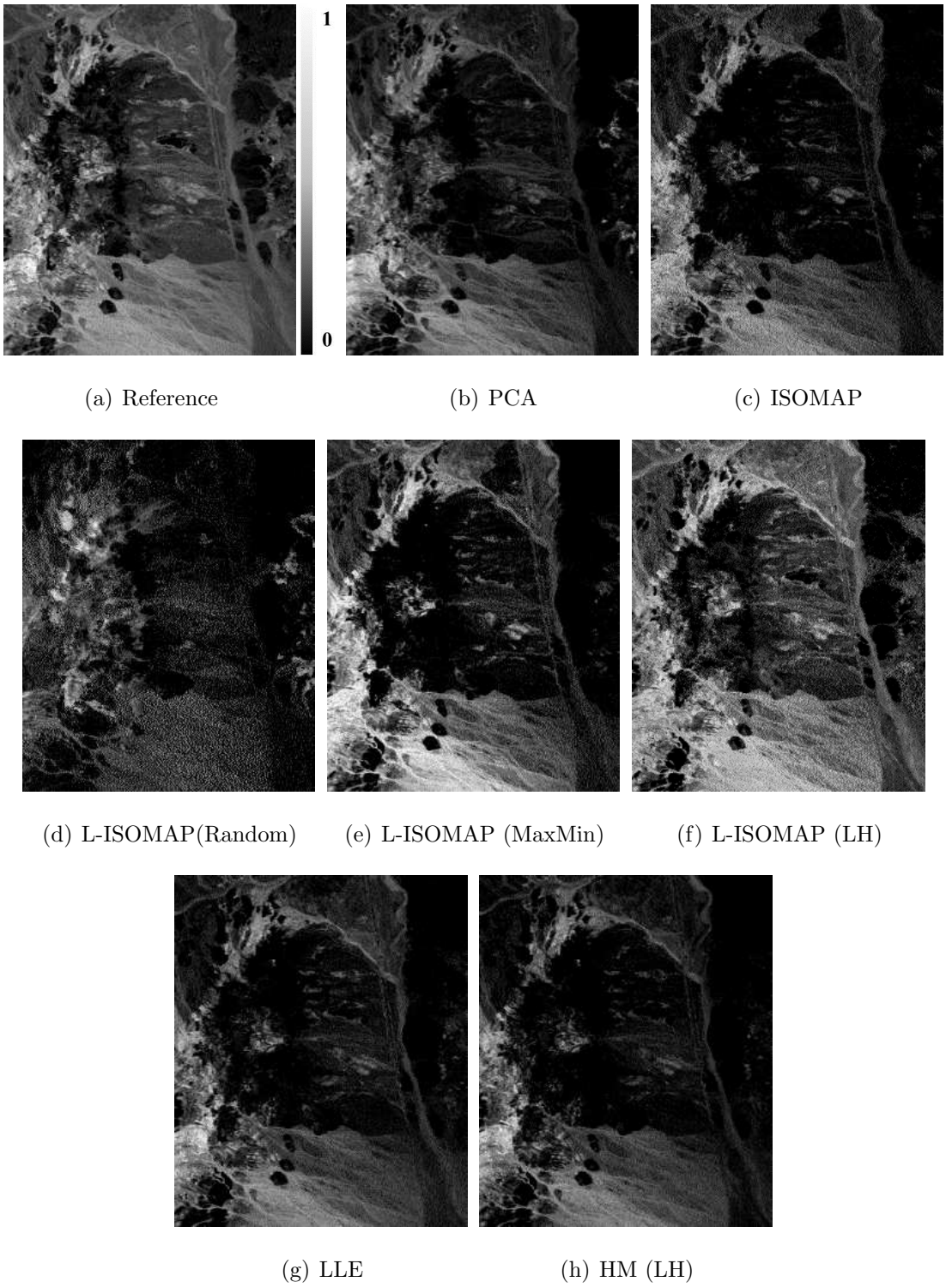


(a) Reference (b) PCA (c) ISOMAP

(d) L-ISOMAP(Random) (e) L-ISOMAP (MaxMin) (f) L-ISOMAP (LH)

(g) LLE (h) HM (LH)

Figure 3.11. Abundance maps of Kaolinite



(a) Reference (b) PCA (c) ISOMAP

(d) L-ISOMAP(Random) (e) L-ISOMAP (MaxMin) (f) L-ISOMAP (LH)

(g) LLE (h) HM (LH)

Figure 3.12. Abundance maps of Muscovite

Table 3.4 Comparison of correlation between extracted abundance map and reference

Correlation	PCA	ISOMAP	L-ISOMAP			LLE	HM (LH)
			Random	MaxMin	LH		
Alunite	**0.9150**	0.8982	0.8030	0.8677	0.7614	0.8902	0.8780
Buddingtonite	0.7816	0.8019	0.7149	0.7284	0.8051	0.6756	**0.8702**
Calcite	0.8860	0.8470	0.8743	**0.8879**	0.8750	0.7274	0.7843
Kaolinite	0.5805	0.7335	0.7361	**0.7790**	0.7295	0.6839	0.7296
Muscovite	0.5103	0.8820	0.9220	0.8973	**0.9579**	0.9175	0.8888
Average	0.7347	**0.8325**	0.8101	0.8320	0.8258	0.7789	0.8302

3.4 Conclusions

In this chapter, a new landmark (backbone point) selection method and a hybrid manifold were investigated for the use of nonlinear manifolds in conjunction with spectral unmixing of hyperspectral remotely sensed data. The proposed methods were compared quantitatively and qualitatively to classical manifold learning methods and other landmark selection approaches. Three conclusions primarily resulted from the experiments:

- The proposed landmark selection method increased both the spectral and spatial diversity of the landmarks compared to the random selection and the maxmin methods. The smaller mean and standard deviation of the SAD from the proposed method were helpful in reducing the variability associated with the extraction process.

- The HM, which shared common characteristics with both the global and local structures of nonlinear manifolds, was advantageous in terms of both unmixing accuracy and processing time.

- The proposed landmark selection method was successfully incorporated in developing the HM.

4. ACTIVE LANDMARK SAMPLING IN MANIFOLD LEARNING BASED UNMIXING

In this chapter, the limitations in the number of proposed landmarks of the proposed landmark selection method described in Chapter 3 are addressed. The resulting landmark set identified by the LH method was relatively small for developing a smooth, stable manifold when the spectral data have complex geometry in the spectral space. A strategy is proposed whereby additional landmarks are selected using an "*active sampling strategy*," based on the initial landmark points.

This chapter introduces an "*active landmark sampling*" framework to select the set of landmarks with the fewest points that are useful, while providing a high quality representation for the spectral unmixing task. To effectively select the landmarks, a simplex volume based ranking method that is implemented on image subsets is proposed [1].

4.1 Active Learning for Supervised Classification of Hyperspectral Data

Humans often learn interactively from past experiences. Interest in how we can "*actively learn*" via interactive feedback has increased because of the capability of these strategies to learn a model more precisely and quickly. Active learning was motivated by scenarios in which it was easy to obtain large quantities of unlabeled data, though difficult to create the labeled data for supervised classification. The goal was to force the learner to effectively learn the target model so that a less costly set with low redundancy could be found [81–83].

[1]The part of this chapter is written based on [80] J. Chi and M. M. Crawford. Active landmark sampling for manifold learning based spectral unmixing. *IEEE Geoscience and Remote Sensing Letters*, which is currently in revision.

In the supervised classification task, defining the efficient training set is critical for obtaining a successful classification result. However, generation of an appropriate training set can be time consuming and expensive. For supervised classification of image data, the training set is typically defined manually in spatial patches and is therefore highly redundant. Further, noisy pixels or anomalies included in the training set may affect calculation of class statistics, which may lead to poor classification accuracy or overfitting [84, 85]. This can result in overuse of a limited number of data points in the classification of remotely sensed data, particularly if training sets are selected manually from the imagery. Active learning, which allows for effective selection of additional data to be labeled based on a small set of previously chosen training data, is growing in popularity for supervised classification [81].

4.2 Active Landmark Sampling

The active learning framework can also be extended to landmark selection methods to improve the representation of the resulting manifold approximations. Although the landmark points identified by the proposed landmark selection approach described in Chapter 3 are spectrally and spatially informative and might be either on boundaries or interior to a class, the number of points is typically smaller than the necessary number of points required to develop a stable, representative manifold for spectral data exhibiting complex geometry. The landmark points identified by the LH method ignore anomalies, but may miss small isolated areas which might contain meaningful information to develop the manifold.

Unlike active learning in classification, where selected data are labeled, equivalent "true" fractional values are not available for the selected pixels, and query criteria must be formulated relative to the metrics for spectral unmixing. The query strategy consists of two phases: 1) ranking batches of candidates from a pool according to some ranking criteria and 2) evaluating the batch of candidate landmarks according to evaluation criteria. Details are included in the next section. Fig. 4.1 illustrates

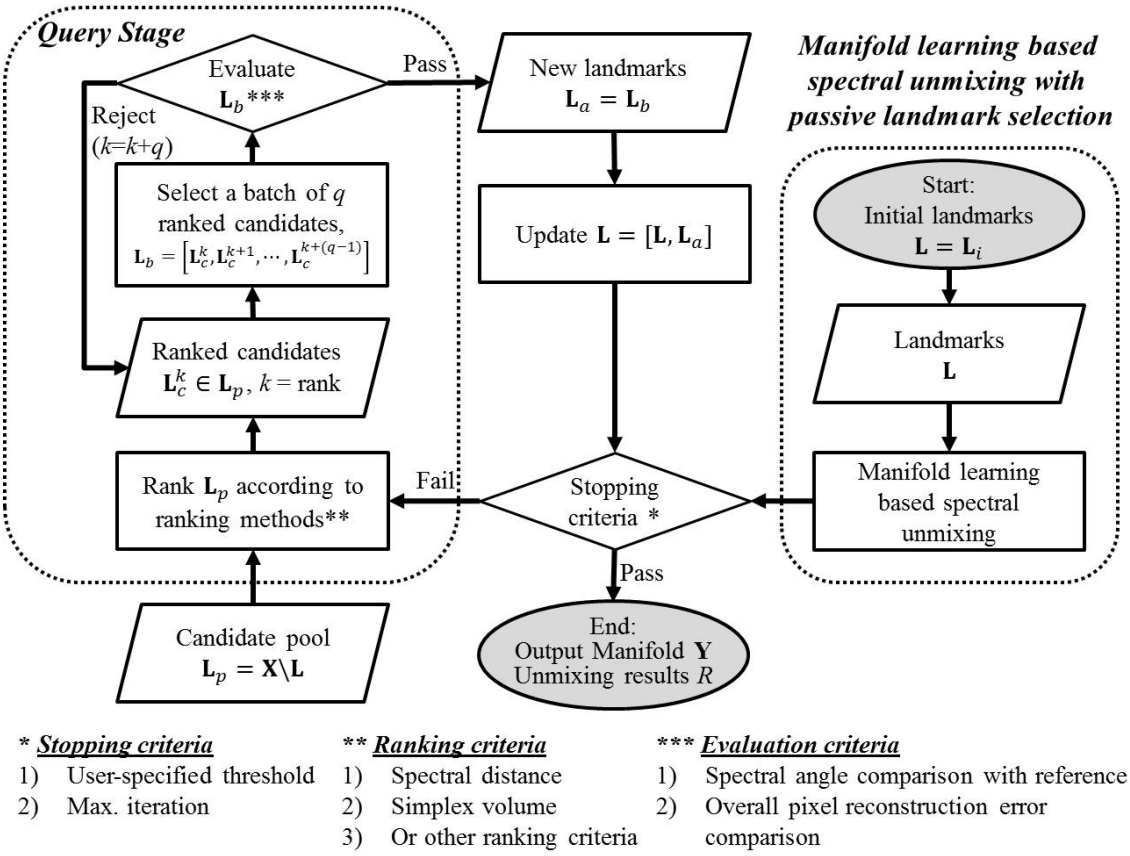


Figure 4.1. Framework of active landmark sampling be selected and the rest

a general framework for the proposed active landmark sampling approach, which is described as follows:

1. Select an initial set of n landmarks $\mathbf{L} = \mathbf{L}_i$ by some landmark selection method.

2. Develop a lower dimensional manifold space $\mathbf{Y}$ using $\mathbf{L}$ and perform spectral unmixing using $\mathbf{Y}$ to obtain an initial unmixing accuracy R.

3. Rank a pool (non-landmark pixels), $\mathbf{L}_p = \mathbf{X} \backslash \mathbf{L}$ according to some ranking criterion, where $\mathbf{X} \in \mathbf{R}^D$ denotes the hyperspectral data.

4. Select a batch of q additional candidate landmarks $\mathbf{L}_b = [\mathbf{L}_c^k, \mathbf{L}_c^{k+1}, \ldots, \mathbf{L}_c^{k+(q-1)}]$ from the ranked $\mathbf{L}_p$, where k is its ranking.

5. Evaluate members of the batch $\mathbf{L}_b$ according to an unmixing based evaluation criterion.

6. If the value of R increases, incorporate the additional landmarks $\mathbf{L}_a = \mathbf{L}_b$ into $\mathbf{L}$. Otherwise, retrieve the next batch of candidates $\mathbf{L}_b = [\mathbf{L}_c^k, \mathbf{L}_c^{k+1}, \ldots, \mathbf{L}_c^{k+(q-1)}]$, where $k = k + q$.

7. Repeat Steps 3-6 until the improvement is less than a user-specified threshold or the maximum number of iterations is reached.

4.3 Query Criteria for Active Learning: Ranking Methods

The actively sampled points can be ranked using various methods. Random sampling is computationally cheaper than any other ranking method since a batch of candidates is directly retrieved from a candidate pool, and may provide high spectral diversity. However, the points are not guaranteed to be sufficient or contribute to improved unmixing accuracy. (note: a random sampling strategy bypasses Step 3 and candidate landmarks are directly retrieved from the pool $\mathbf{L}_p$.)

4.3.1 Spectral Distance Based Ranking of Candidate Landmarks

Diversity is an important condition for selecting good landmarks. The spectral distance to the initial landmark points is straightforward for estimating the diversity of the landmarks. The pixels with the maximum mean spectral distance from the initial landmarks may increase the diversity, but may also include noise or outliers. This might result in manifolds that are not smooth or have sub-manifold structures if the data are sparse. Alternatively, if the landmarks are selected from the pixels that are spectrally close to the initial set, they might represent locally complex geometry, although small isolated areas may be ranked lower and may result in spectral redundancy. In this study, we initially rank all the pixels in the candidate pool $\mathbf{L}_p$ according to Euclidean mean distance to the initial landmark set, and then evaluate

a batch of q candidates from the pixels with the maximum or minimum distance to the $\mathbf{L}$.

4.3.2 Sequential Simplex Volume Ranking Method

In the active landmark sampling framework described in Fig. 4.1, the candidate pool $\mathbf{L}_p$ should be re-ranked when $\mathbf{L}$ is updated at each learning step. Defining $\mathbf{L}_p$ as all the pixels that are not in $\mathbf{L}$, can result in spectral redundancy in the pixels. While pixels may not have spectrally distinguishing characteristics relative to the full image, they may be spectrally unique relative to a spatially local area. A new ranking method is proposed that first seeks to increase the chance of detection of locally high spectral variability using sequentially smaller image subsets, and then ranks the pixels by the volume of the simplex spanned by the previously selected landmarks $\mathbf{L}$ and a batch of candidate points $\mathbf{L}_b$ retrieved from the candidate pool.

The proposed method, "*sequential simplex volume (SSV)*," is composed of three phases: 1) performing singular value decomposition (SVD) to determine a set of eigenvectors of image subsets; 2) determining landmark candidates by projecting the image data onto the eigenvectors; and 3) ranking the candidates according to the magnitude of the simplex volume formed by the previously selected landmark set and the candidates. This method starts with a full image and then successively divides it into a smaller subsets to detect pixels that contribute to local spectral variability. Use of spatial subsets is computationally advantageous and results in landmarks that are potentially well distributed across the image.

Step1. Eigenvector determination

SVD is a very efficient projection technique commonly used in remote sensing to obtain a set of eigenvectors that explain most of the spectral variability of the data [86, 87]. In hyperspectral data with high inter-band spectral correlation, PCA may fail if subtle differences occur in individual bands and the optimal feature space

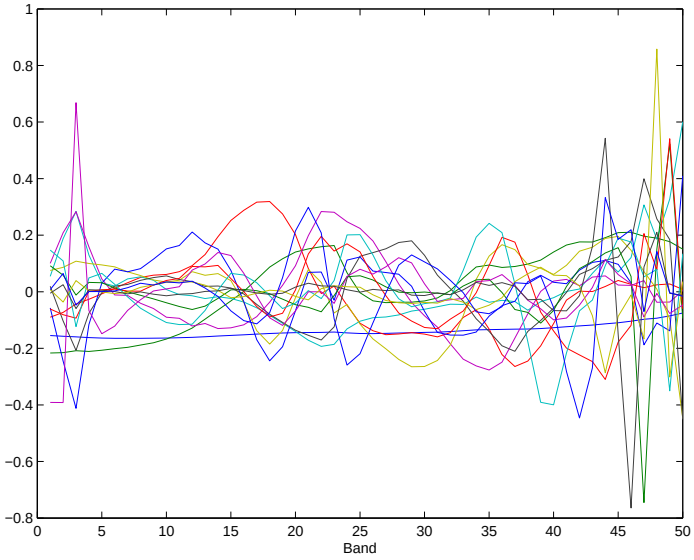


Figure 4.2. Eigenvectors from an example image subset

varies with different classes in the data [60]. MNF requires a noise covariance matrix that is difficult to obtain from small subsets. SVD can address the "singularity" problem related spectral correlation in hyperspectral data [54, 88]. The SVD of $[\mathbf{X} : \mathbf{x}_i, i = 1, \ldots, N] \in \mathbf{R}^D$ is defined by [89]

$$\mathbf{X} = \mathbf{USV}^T \tag{4.1}$$

where $\mathbf{V}^T$ contains the unit row eigenvectors of $\mathbf{X}^T\mathbf{X}$ sorted in order of descending significance, $\mathbf{S}$ is of the form $\left(\begin{smallmatrix} \mathbf{C} & 0 \\ 0 & 0 \end{smallmatrix}\right)$, $\mathbf{C}$ is diagonal, the square roots of the eigenvalues of $\mathbf{X}^T\mathbf{X}$ are sorted in descending value, and $\mathbf{U}$ denotes a matrix containing the unit column eigenvectors of $\mathbf{XX}^T$.

The SSV method uses sequential, nonoverlapping, equal sized subsets of the image. As shown in Fig. 4.2, for each subset, the eigenvectors accounting for 99% of the total spectral variance are retained from each subset and then compiled into a single eigenvector matrix $\mathbf{E}$. The vectors represent high spectral variability candidates in the local spatial patches. From larger subsets, candidates might have relatively low spectral variability, indicating that they are useful for large homogeneous regions. The

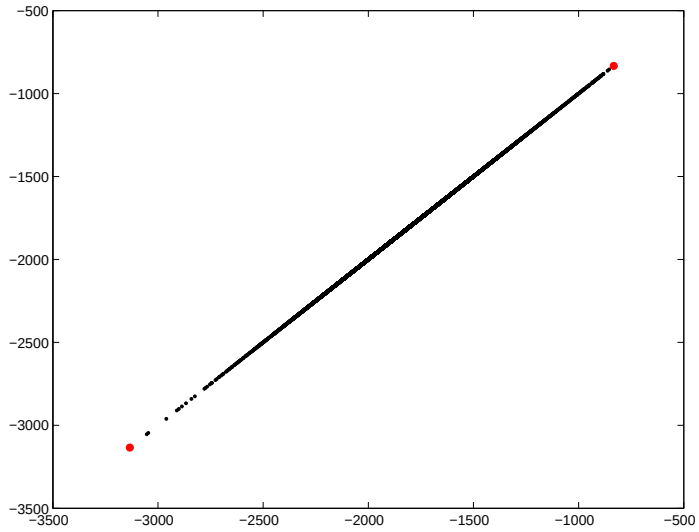


Figure 4.3. Projection of data points onto one of eigenvectors

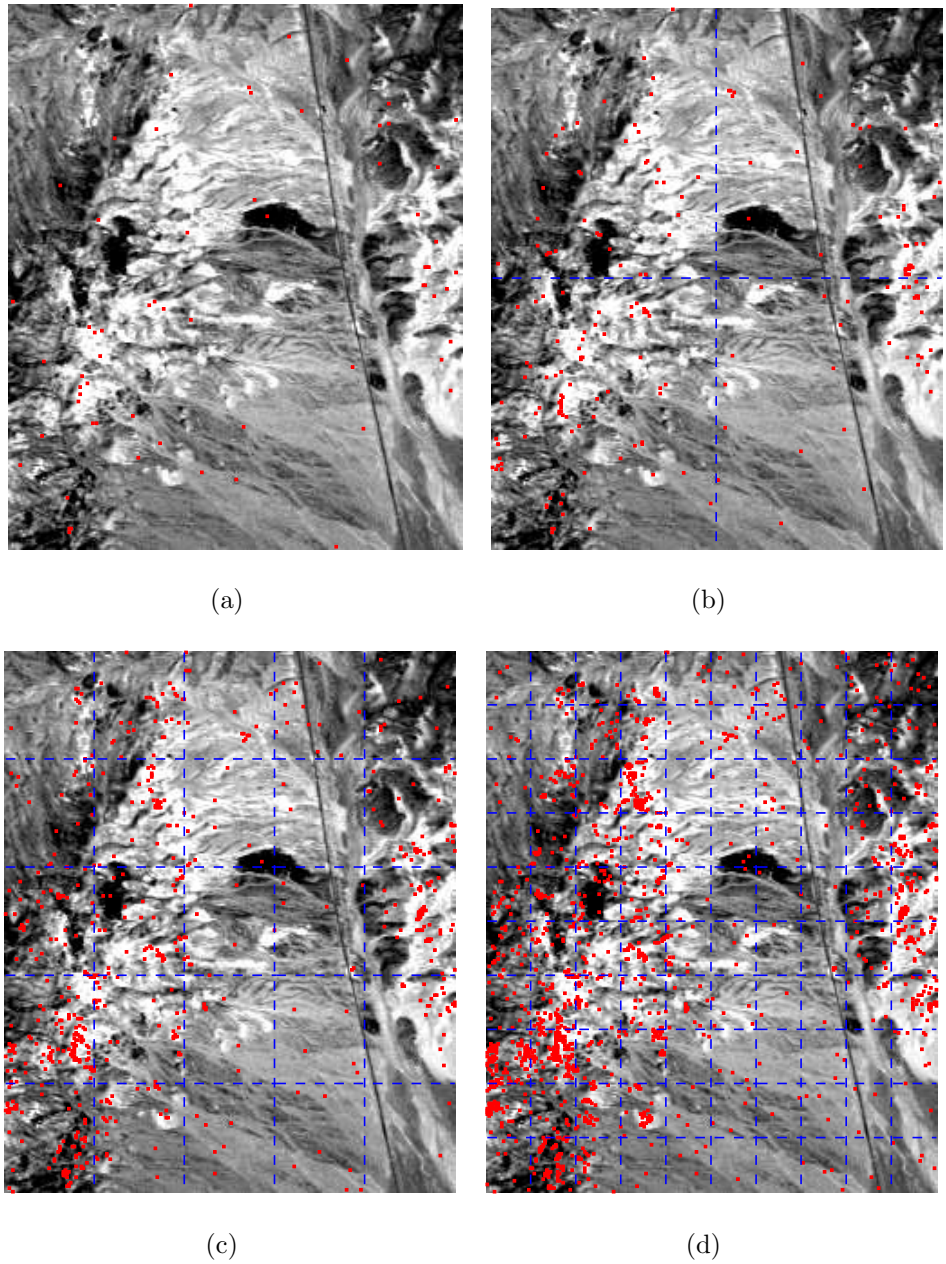
use of smaller subsets might obtain more representative candidates for local spectrally complex regions, although a large number of eigenvectors is retained ultimately.

Step2. Projection

Fig. 4.3 illustrates the projection of the entire set of data points onto one of the eigenvectors $\mathbf{E}$ obtained in Step 1. The projection can be expressed by the following:

$$\mathbf{P} = \mathbf{E}^T\mathbf{X} \tag{4.2}$$

where $\mathbf{E}$ is comprised of the eigenvectors, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$ is the full data matrix, and $\mathbf{P}$ is the projected data matrix. The pixels lying at either extreme (e.g., red dots in Fig. 4.3) of the projection are more likely to be good candidates, and are identified from each subset. As the number of pixels of the subsets decreases, however, the number of candidate pixels increases, which indicates that the spectral redundancy would also increase (see Fig. 4.4). To address this problem, identifying the representative pixels by the geometric connectivity used in the LH method is also implemented to remove the spectral redundancy between pixels that are connected on the spatial graph.



(a) (b)

(c) (d)

Figure 4.4. Candidates from different size of subsets: (a) 300×250, (b) 150×125, (c) 60×50, (d) 30×25

Step3. Ranking by simplex volume

Similar to convex geometry-based endmember extraction algorithms, landmarks forming a simplex can be used for manifold approximations to embed the remaining

data points. The volume of the simplex spanned by $\mathbf{L}$ and $\mathbf{L}_b$ is computed to rank the candidates. However, the simplex volume computation (3.9) used in N-FINDR is violated because the number of features is much smaller than the number of landmarks, and the landmark set should be updated to be associated with the active sampling framework. For these reasons, the simplex volume calculation is reformulated using the Gram determinant.

The volume of the simplex with $n+1$ landmarks $[\mathbf{L}, \mathbf{L}_b] = [\mathbf{l}_0, \mathbf{l}_1, \ldots, \mathbf{l}_n]$, is

$$V(\hat{\mathbf{M}}) = \frac{1}{n!} \operatorname{abs}(\det(\hat{\mathbf{M}})) \tag{4.3}$$

where $\hat{\mathbf{M}} = [\mathbf{l}_1 - \mathbf{l}_0, \mathbf{l}_2 - \mathbf{l}_0, \ldots, \mathbf{l}_n - \mathbf{l}_0]$, are the vertices of the n-simplex formed by the origin along with these vertices (i.e., shifting the origin to $\mathbf{l}_0$). The value of the matrix determinant is not changed by adding a multiple of one row to another row, or a multiple of one column to another column; $\mathbf{M}$ in (3.9) and $\hat{\mathbf{M}}$ in (4.3) have the same determinant. So, (4.3) is equivalent to (3.9) in simplex volume computation. Since $\left|(\det(\hat{\mathbf{M}})\right| = (\det(\hat{\mathbf{M}}^T\hat{\mathbf{M}}))^{1/2}$, $\det(\hat{\mathbf{M}})$ can be calculated by the Gram determinant, which is based on a Gram-Schmidt orthogonalization process as follows [90, 91]:

$$\det(\hat{\mathbf{M}}) = \begin{bmatrix} \mathbf{a}_1^T\mathbf{a}_1 & \mathbf{a}_1^T\mathbf{a}_2 & \cdots & \mathbf{a}_1^T\mathbf{a}_n \\ \mathbf{a}_2^T\mathbf{a}_1 & \mathbf{a}_2^T\mathbf{a}_2 & \cdots & \mathbf{a}_2^T\mathbf{a}_n \\ \cdots & \cdots & \ddots & \cdots \\ \mathbf{a}_n^T\mathbf{a}_1 & \mathbf{a}_n^T\mathbf{a}_2 & \cdots & \mathbf{a}_n^T\mathbf{a}_n \end{bmatrix}^{1/2} \tag{4.4}$$

where $\mathbf{a}_1 = \mathbf{l}_1 - \mathbf{l}_0, \mathbf{a}_2 = \mathbf{l}_2 - \mathbf{l}_0, \ldots, \mathbf{a}_n = \mathbf{l}_n - \mathbf{l}_0$. A simple method to calculate the Gram determinant is based on orthogonal vectors. Let $[\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \ldots, \hat{\mathbf{a}}_n]$ be the corresponding vectors of $[\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$ after the Gram-Schmidt orthogonalization process. Then (4.4) becomes

$$\det(\hat{\mathbf{M}}) = \begin{bmatrix} |\hat{\mathbf{a}}_1|^2 & 0 & \cdots & 0 \\ 0 & |\hat{\mathbf{a}}_2|^2 & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & |\hat{\mathbf{a}}_n|^2 \end{bmatrix}^{1/2} = |\hat{\mathbf{a}}_1||\hat{\mathbf{a}}_2|\cdots|\hat{\mathbf{a}}_n| \tag{4.5}$$

While (3.8) must have n dimensional data to extract n+1 landmarks, the Gram determinant does not have this limitation since $\hat{\mathbf{M}}$ is always a square matrix.

Once the candidates are ranked according to their volumes, they are further evaluated by additional query criteria. Some points (e.g., outliers) might not be useful for developing more representative manifolds (e.g., the pixel in LL of Fig. 4.3), and some (e.g., small isolated areas) might be used as a new landmark point (e.g., the pixel in UR of Fig. 4.3). If the candidate pool $\mathbf{L}_p$ is empty, then Steps 1-3 are repeated using smaller subsets to examine the candidates that have locally high "contrast".

The LH method for selecting the initial landmark set is based on spectral reflectances and uses the SAD to compute the average distance between pixels in the kernel. It ignores small isolated areas that may include interesting features due to smoothing effects by the window kernel. The proposed SSV method, however, utilizes the most spectrally meaningful pixels in the transformed feature domain in nonoverlapping spatial subsets of data. The SSV method might capture the inherent features from the sequential SVD components that are overlooked in the LH landmark selection method. Additionally, since it exploits convex geometry to rank the landmark candidates, the idea is consistent with N-FINDR to find the endmembers in the unmixing phase.

4.4 Query Criteria for Active Learning: Evaluation Methods

A batch of q ranked candidates does not always guarantee better landmarks for manifold approximations. Although they are highly ranked by a ranking criterion such as spectral distance or simplex volume, they may not be useful to develop a

representative manifold if they have been selected from anomalous pixels. For these reasons, an evaluation stage should be implemented in the active landmark sampling framework to determine whether the batch of pixels should be accepted or not.

Two evaluation criteria are used for active learning related to landmark selection. If reference information is available, comparison of spectral angle distance (SAD) with the ground reference can be used to evaluate the quality of each endmember extracted from the image. However, in practice, it is often difficult to obtain reference data. In these cases, statistics such as RMSE are often used to evaluate the overall difference between the original and the reconstructed image using fractional abundance maps, which are described as follows:

$$\mathbf{RMSE} = \sqrt{\frac{1}{D}\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_i - \mathbf{y}_i\|^2} \tag{4.6}$$

$$\mathbf{y}_i = \sum_{j=1}^{q} a_{i,j} \times \mathbf{e}_j \tag{4.7}$$

where $\mathbf{x}_i$ and $\mathbf{y}_i$ denote the original hyperspectral signature and reconstructed pixel vector using fractional abundances $a_{i,j}$ of endmember $\mathbf{e}_j$, respectively.

4.5 Experimental Results

Experiments were conducted to evaluate the proposed active landmark sampling strategy (AL) and sequential simplex volume-based ranking method. The same AVIRIS Cuprite dataset was used, and results were compared to those obtained using the hybrid manifold (HM) method implemented with the LH approach in Chapter 3. The candidate points in the active landmark sampling framework were ranked via random sampling, maximum (MaxDist) and minimum (MinDist) spectral distance based ranking methods, and the proposed SSV method. The initial 324 landmarks for the active sampling were identified using the LH method, and a batch of 3 candidate pixels was selected from the ranked candidate pool at every learning step.

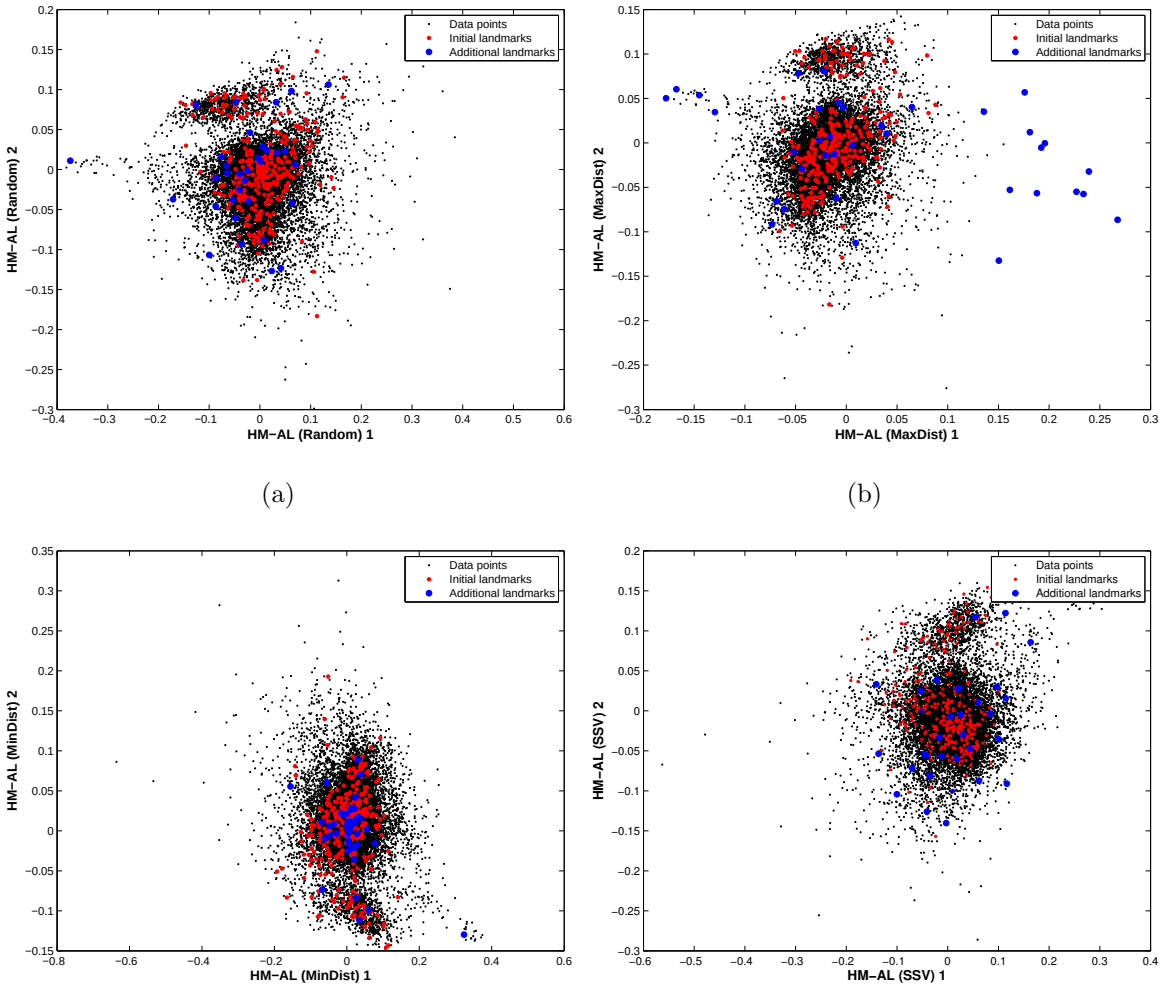


Figure 4.5. Distribution of extracted landmarks by active sampling using: (a) random selection, (b) maximum distance, (c) minimum distance, and (d) sequential simplex volume ranking methods

4.5.1 Distribution of Landmark Points

Fig. 4.5 shows distribution of the initial landmarks across the image and the additional actively sampled landmark points identified by different sampling methods. Similar to the results in the previous chapter, the additional landmarks obtained by the random sampling are uniformly distributed within the scatter plot (Fig. 4.5(a)), but they induce uncertainty into the results and guarantee neither spectrally nor

spatially meaningful information. Since the additional landmarks identified by the MaxDist ranking method are the pixels that are spectrally furthest from the initial points, most pixels were found in the outer fringe of the data clouds (see Fig. 4.5(b)). So, this resulted in development of relatively sparse manifold coordinates although the spectral diversity of the landmarks increased. As shown in Fig. 4.5(c), the spectral diversity of the landmarks by the MinDist was low compared to that of the random and the MaxDist methods, but more points were selected from very dense point clouds in the middle of the scatter plot, which may have complex local geometry. The resulting manifold may be more stable than than the manifolds associated with other ranking methods, but the points by the MinDist method are concentrated in a narrow range of values compared to the MaxDist. The proposed method selected the additional landmarks primarily from the regions which might lack landmarks, but not from the very small areas which might be outliers (see Fig. 4.5(d)). However, some were selected near the initial set, which may be improving the approximation related to complex geometry in a given manifold space.

4.5.2 Evaluation of Endmembers via Spectral Angle Distance

The manifolds, ranking and evaluation criteria are compared in terms of spectral angle distance (SAD) as listed in Table 4.1. The average SADs of the HM-AL methods were lower than the other manifold approaches. In particular, kaolinite and muscovite, which are difficult endmembers because of the similar absorption features, had lower SAD than the other manifolds. Since active landmark sampling for spectral unmixing collects additional landmarks until there is no further incremental improvement, the difference in the error among the AL ranking methods is not significant. All the methods eventually converged to a high accuracy outcome. However, as shown in Table 4.1 and Fig. 4.6, random sampling required about 200 learning steps to converge to the average SAD value achieved by the other ranking methods, although the error decreased rapidly during the first learning steps. The additional

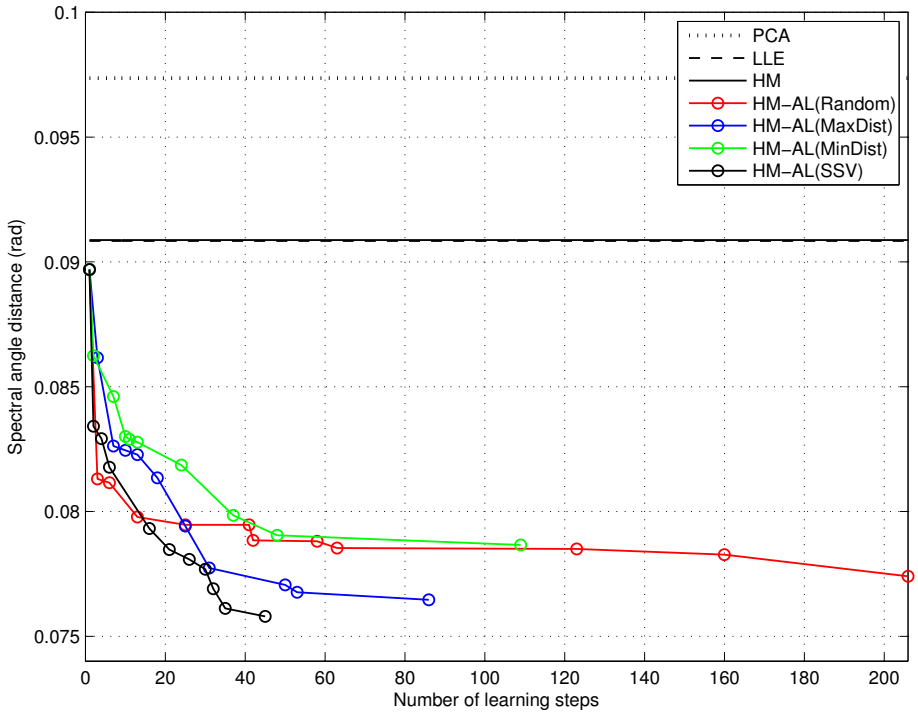


Figure 4.6. Changes of average spectral angle distance of five mineral endmembers as a function of the number of learning steps

landmarks by other ranking methods did not contribute quite as much, indicating that the original samples might have been too sparse over the full manifold space. The MaxDist criterion did not significantly reduce the error initially since it might have sought additional points from the outlier. However, once some landmarks were selected from the data clouds, the error rapidly decreased. Although initial error decreased slowly when the MinDist criterion was used, the error at convergence was the same as for other methods. The proposed method found the spectrally and spatially meaningful landmarks via nonoverlapping spatial subsets of decreasing size as shown in Fig. 4.4 and 4.5(d). More accurate results were achieved with fewer learning steps than the other ranking methods since the landmarks were not spectrally or spatially redundant. Additionally, the points covered parts of the manifold that might not be well represented in the other methods.

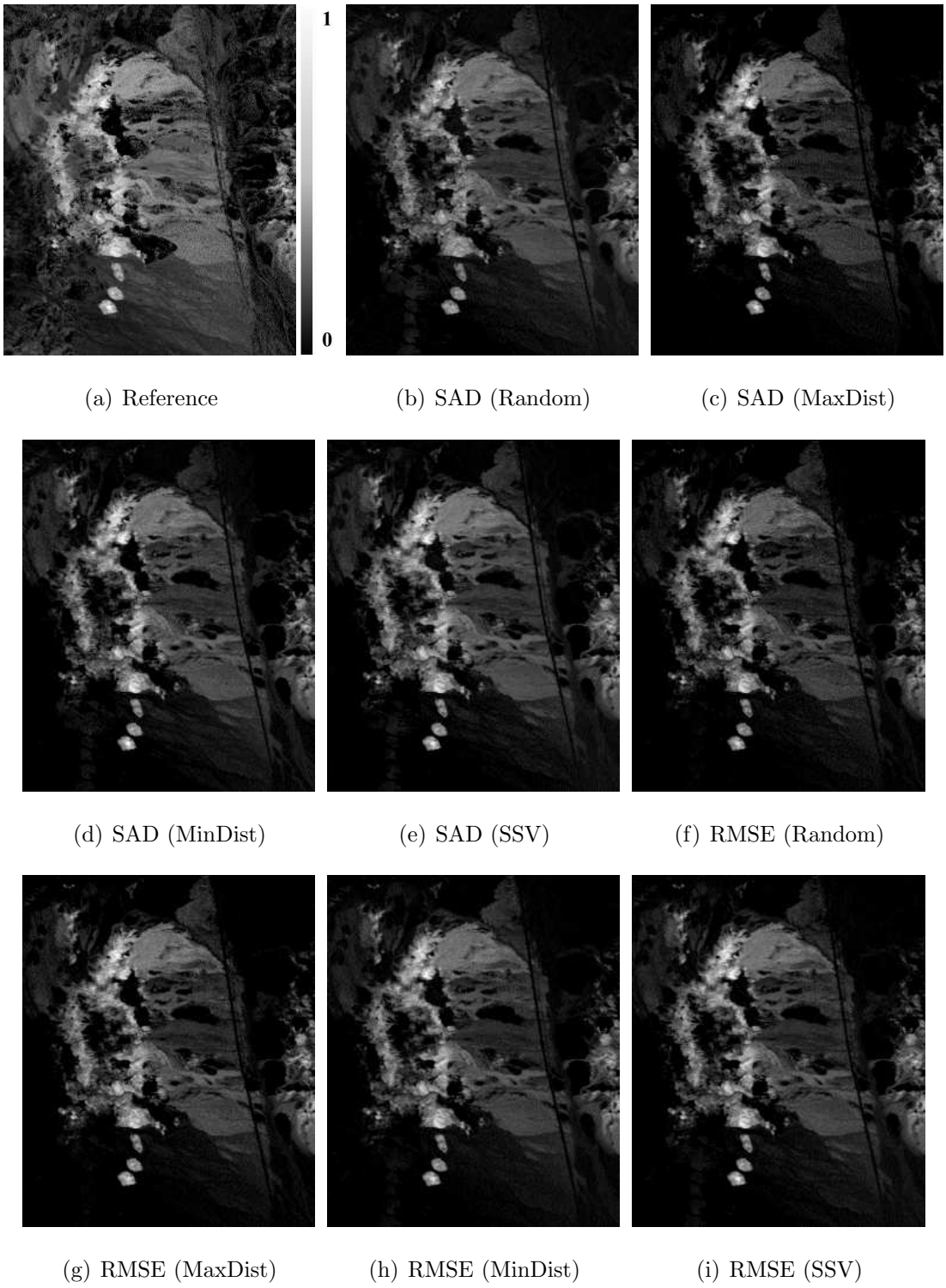
When the reconstruction error for the evaluation criterion was used, all ranking methods except random sampling converged with low reconstruction errors, but required more queries than those for the SAD-based methods. The mean values of the SAD were generally higher than the SAD-based evaluation methods, as shown, but lower than the other approaches without active learning. Thus, while the reconstruction error is worthwhile to utilize as an alternative metric if ground reference information is not available, the reconstruction error may not always guarantee better unmixing quality.

4.5.3 Evaluation of Endmembers via Fractional Abundance Map

The fractional abundance maps were evaluated in the same way as in the previous chapter. Similar to SAD based comparisons, the average correlation between extracted abundance map by HM-AL methods and the reference was higher than the values obtained by methods that did not incorporate an active sampling strategy, as indicated by the values in Table 4.2. The HM-AL methods found more accurate endmembers than other manifold approaches, and had visually higher agreements with the reference abundance maps, as shown in Fig. 4.7-4.11. Similar to SAD and correlation comparisons, it is difficult to compare the abundance maps visually since the results are very similar. For buddingtonite (Fig. 4.8) and muscovite (Fig. 4.11), however, the abundance maps by the proposed SSV method had better agreement with the reference information, while others (Fig. 4.7, 4.9, 4.10) generally underestimated the fractional abundances. When the reconstruction error was used as the criterion, random and spectral based sampling methods had slightly lower correlations with the reference than the baseline method (see Table 4.2) although they resulted in higher accuracy in SAD comparisons (see Table 4.1). However, the proposed SSV ranking method produced higher correlations than the other methods for both the SAD and RMSE based query criteria for this testbed of data.

Table 4.1 Comparison of spectral angle distance of hybrid manifold learning with active landmark sampling

Spectral Angle	HM (LH)	HM-AL							
Evaluation method		Spectral Angle				Reconstruction Error			
Ranking method		Random	MaxDist	MinDist	SSV	Random	MaxDist	MinDist	SSV
Alunite	0.0678	0.0614	0.0513	0.0500	**0.0499**	0.0611	**0.0535**	0.0611	0.0650
Buddingtonite	0.1434	**0.1298**	0.1350	0.1389	0.1331	0.1384	0.1360	0.1395	**0.1331**
Calcite	0.1023	0.0691	**0.0668**	**0.0668**	**0.0668**	**0.0668**	**0.0668**	0.1023	0.0749
Kaolinite	0.0613	0.0567	**0.0530**	0.0613	**0.0530**	0.0613	0.0613	0.0613	**0.0598**
Muscovite	0.0796	**0.0700**	0.0763	0.0763	0.0763	**0.0796**	0.0817	0.0808	0.0817
Average SAD	0.0909	0.0774	0.0765	0.0787	**0.0758**	0.0814	**0.0799**	0.0890	0.0829
Reconstruction Error						**0.0113**	0.0115	0.0114	0.0114
# of learning steps		206	86	109	**45**	170	114	142	**84**

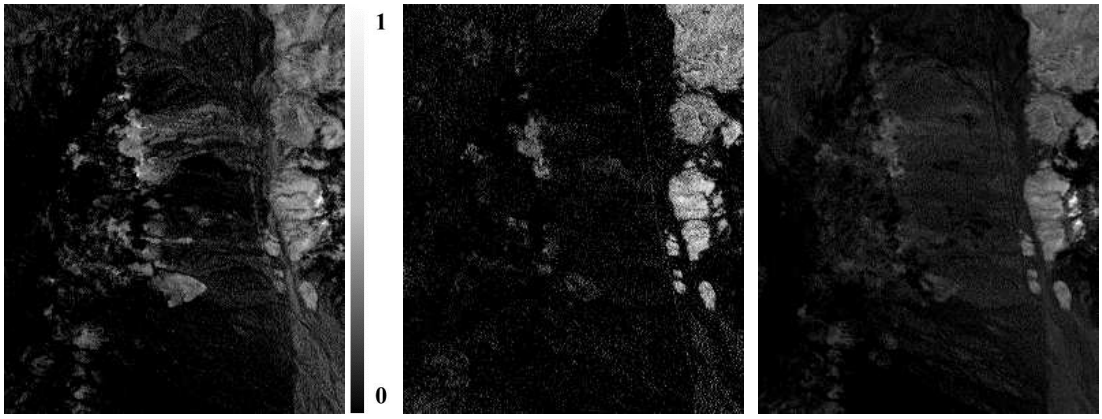


(a) Reference (b) SAD (Random) (c) SAD (MaxDist)

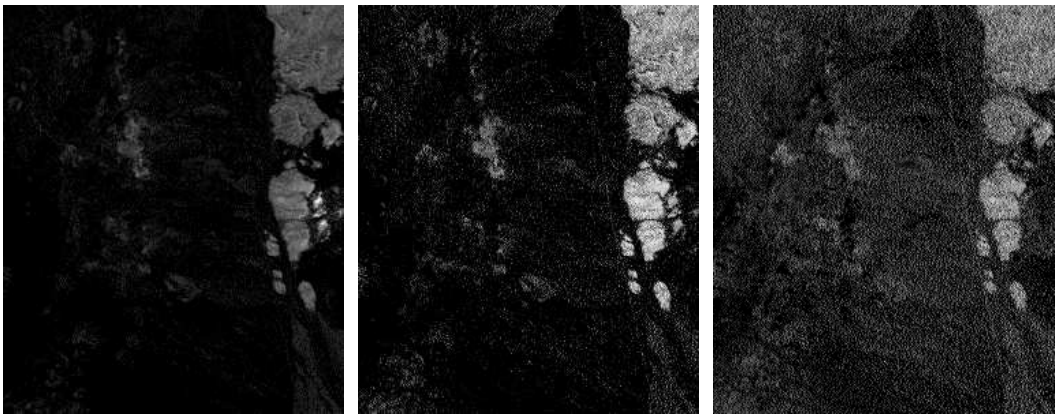
(d) SAD (MinDist) (e) SAD (SSV) (f) RMSE (Random)

(g) RMSE (MaxDist) (h) RMSE (MinDist) (i) RMSE (SSV)

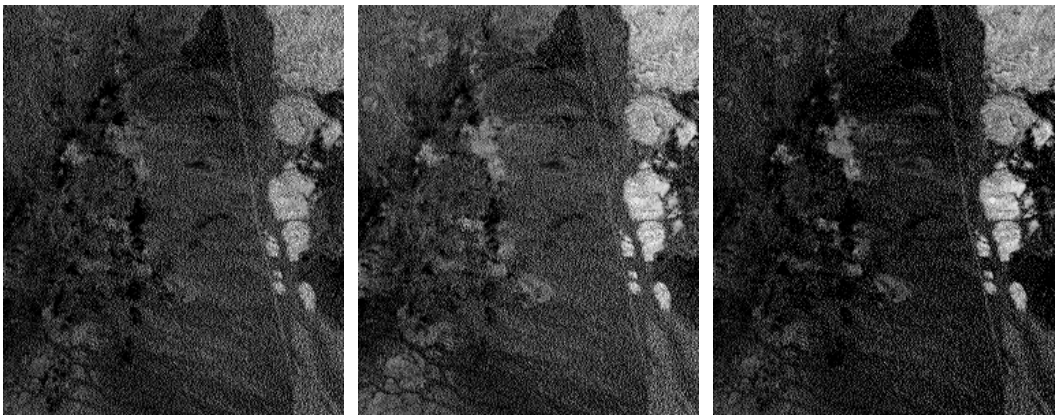
Figure 4.7. Abundance maps of Alunite by Hybrid Manifold with active learning



(a) Reference (b) SAD (Random) (c) SAD (MaxDist)

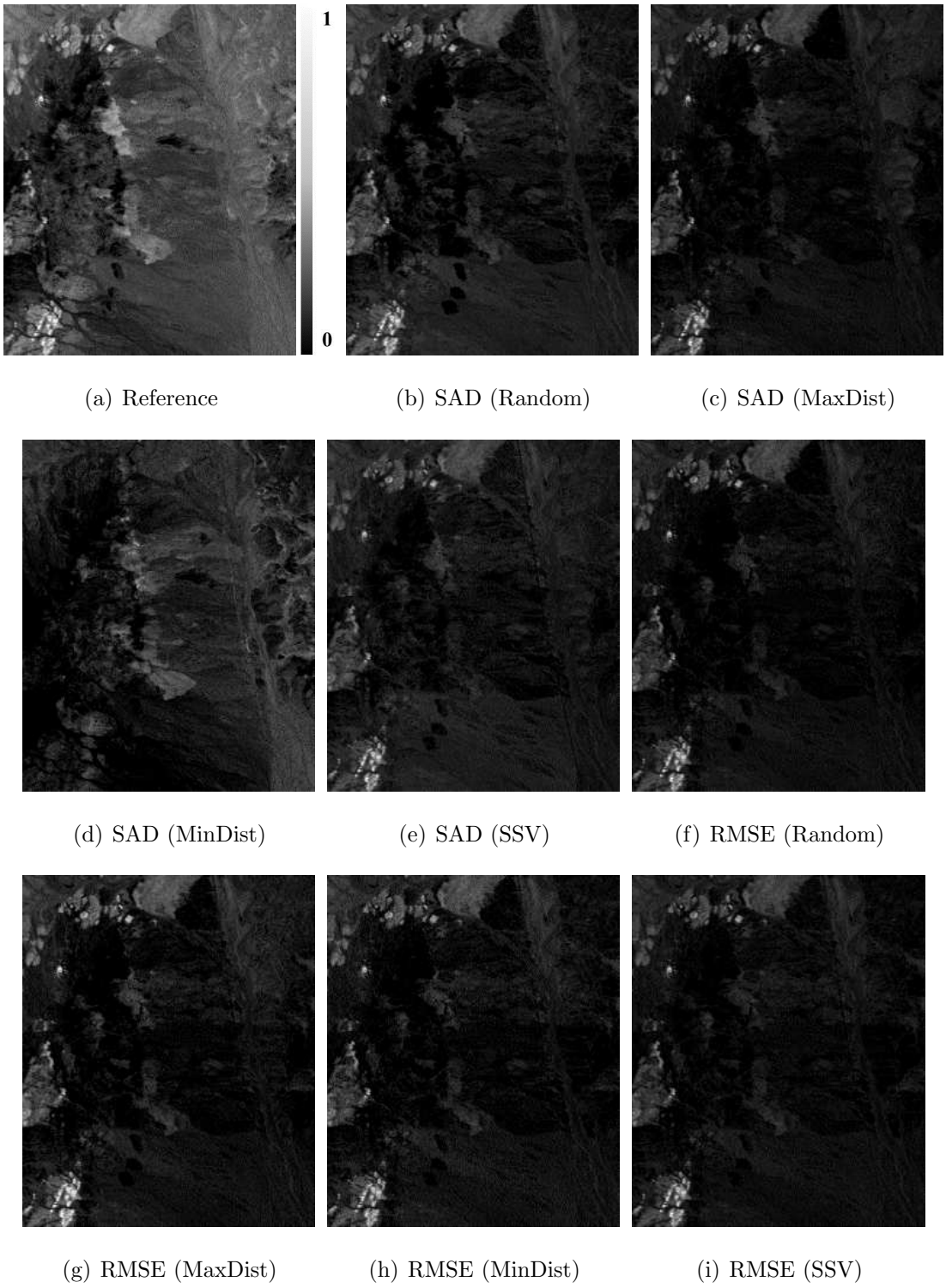


(d) SAD (MinDist) (e) SAD (SSV) (f) RMSE (Random)



(g) RMSE (MaxDist) (h) RMSE (MinDist) (i) RMSE (SSV)

Figure 4.8. Abundance maps of Buddingtonite by Hybrid Manifold with active learning

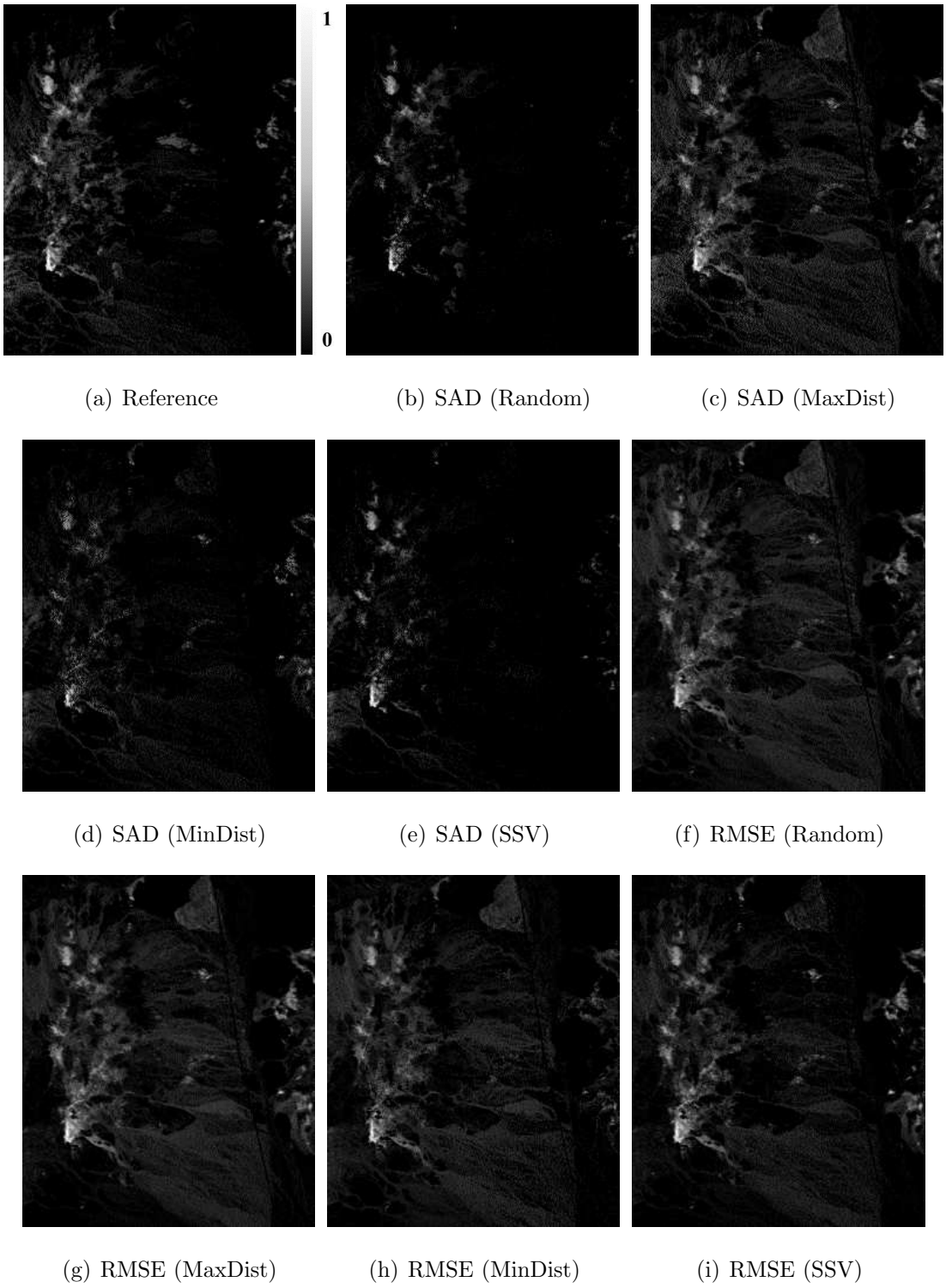


(a) Reference (b) SAD (Random) (c) SAD (MaxDist)

(d) SAD (MinDist) (e) SAD (SSV) (f) RMSE (Random)

(g) RMSE (MaxDist) (h) RMSE (MinDist) (i) RMSE (SSV)

Figure 4.9. Abundance maps of Calcite by Hybrid Manifold with active learning

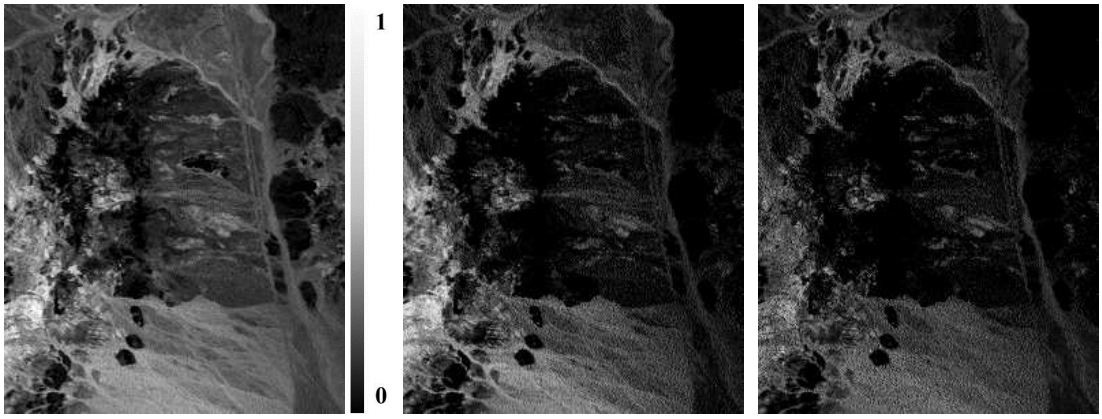


(a) Reference (b) SAD (Random) (c) SAD (MaxDist)

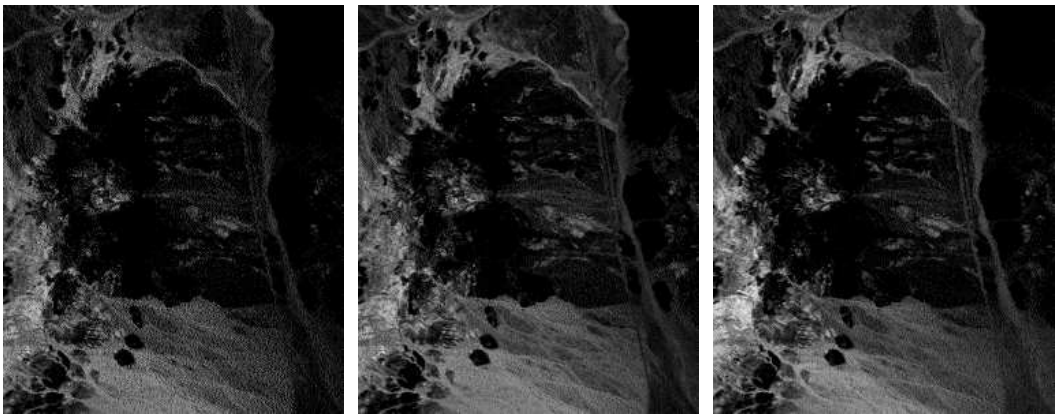
(d) SAD (MinDist) (e) SAD (SSV) (f) RMSE (Random)

(g) RMSE (MaxDist) (h) RMSE (MinDist) (i) RMSE (SSV)

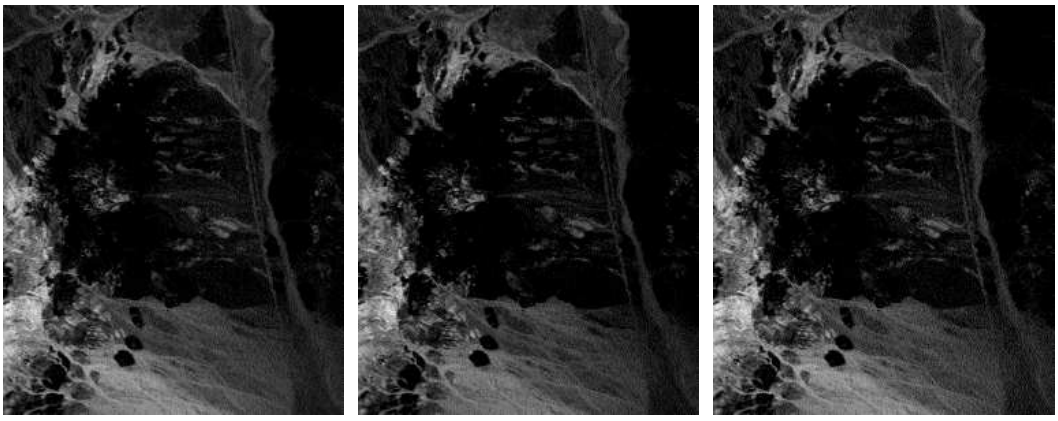
Figure 4.10. Abundance maps of Kaolinite by Hybrid Manifold with active learning



(a) Reference (b) SAD (Random) (c) SAD (MaxDist)



(d) SAD (MinDist) (e) SAD (SSV) (f) RMSE (Random)



(g) RMSE (MaxDist) (h) RMSE (MinDist) (i) RMSE (SSV)

Figure 4.11. Abundance maps of Muscovite by Hybrid Manifold with active learning

Table 4.2 Comparison of correlation between extracted abundance map by Hybrid manifold with active landmark sampling and reference

Correlation	HM (LH)	HM-AL							
Evaluation method		Spectral Angle				Reconstruction Error			
Ranking method		Random	MaxDist	MinDist	SSV	Random	MaxDist	MinDist	SSV
Alunite	0.8780	**0.9147**	0.8611	0.9040	0.9008	0.8714	0.8791	0.8922	**0.8927**
Buddingtonite	**0.8702**	0.7911	0.8645	0.8275	0.8122	0.7921	0.7987	0.8208	**0.8385**
Calcite	0.7843	0.8825	0.8604	0.8839	**0.8897**	0.8105	**0.8218**	0.8159	0.8196
Kaolinite	0.7296	0.7512	0.7716	0.8007	**0.8056**	0.7059	0.6970	0.7097	**0.7535**
Muscovite	0.8888	0.9269	0.9099	0.8992	**0.9473**	**0.9233**	0.9145	0.9065	0.9145
Average Correlation	0.8302	0.8533	0.8535	0.8631	**0.8711**	0.8206	0.8222	0.8290	**0.8438**

4.6 Conclusions

In this chapter, a new active landmark sampling framework and sequential simple volume based ranking method were investigated in conjunction with spectral unmixing of hyperspectral remotely sensed data. The two contributions include:

- The number of initial landmarks identified by the LH method for the HM is quite small. Additional landmarks were selected via an active sampling method, resulting in improved unmixing results relative to the manifolds developed using the original landmarks. The proposed framework was implemented with two query phases which are related to ranking and evaluation methods for spectral unmixing. The points were retained or rejected according to an evaluation criterion.

- The ranking methods all yielded similar average spectral distance or reconstruction error according to the evaluation methods, but the proposed SSV ranking method achieved these results with fewer learning steps than other ranking methods. Since the new ranking method exploited landmark candidates based on feature-level measurement from sequential subsets and convex geometry for ranking, it was incorporated into the active landmark sampling framework for spectral unmixing. Thus, the proposed method could increase both the spectral and spatial diversity of the additional landmarks.

5. CROP RESIDUE ESTIMATION USING UNMIXING TECHNIQUES

Crop residue helps to moderate soil temperature and increase water use efficiency in the short term while also providing improvement in soil quality, increase in soil organic carbon, and facilitation of the biodegradation of pollutants for long term sustainability. Since good management of crop residues can also increase irrigation efficiency and erosion control, use of remote sensing techniques is receiving increased attention for the management of crop residue amounts. Remote sensing index-based methods, but low SNR image data such as Hyperion are challenging for application of these approaches. In this chapter, a spectral unmixing technique is proposed as an alternative approach to the index-based methods to effectively estimate and monitor crop residue cover with airborne and spaceborne hyperspectral imagery [1].

5.1 Background in Crop Residue Study

Crop residues remaining in agricultural fields after the harvest play an important role in controlling and protecting against water and wind erosion, and increasing soil organic carbon. For these reasons, it is important to determine the quantity of residue remaining between the end of the growing season and the onset of the planting season the following year. However, the current approaches for quantifying residue cover are inadequate for characterizing residue amounts over extended areas. Remotely sensed data have the capability to efficiently record reflected energy over

[1]This chapter is extracted from [92] J. Chi and M. M. Crawford. Spectral unmixing based crop residue estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, which is currently in revision.

extended and inaccessible areas using sensors on aircraft or spacecraft, while also providing information on the spatial variability of reflectance [93–97].

Over the past decade, many studies have examined methods for estimating crop residue cover via classification and spectral unmixing, as well as empirical models based on multispectral and hyperspectral remote sensing indices [93–100]. While classification methods categorize pixels into discrete, non-overlapping thematic categories, unmixing techniques provide continuous values that are physically meaningful. Bannari et al. compared the capabilities of hyperspectral and multispectral sensors to estimate and map crop residue, soil, and crop covers using linear spectral mixture analysis [95]. The hyperspectral data captured the absorption features of the chemistry related to the amount of residue cover much better, and outperformed the multispectral sensor for estimating residue cover. In [96], Pacheco and McNairn demonstrated that the spectral unmixing technique applied to multispectral Landsat and SPOT imagery to produce crop residue estimates in terms of multi-temporal analysis and residue types.

Remote sensing indices such as the Normalized Difference Vegetation Index (NDVI) [101], the ratio of the difference between the reflectances in the NIR and the red region of the spectrum and their sum, has often been utilized for representing the relative greenness in vegetation. Image derived indices have been utilized as independent variables in regression models to estimate the quantity of ground materials. In residue research, the Normalized Difference Tillage Index (NDTI) [93] and Cellulose Absorption Index (CAI) [94, 98, 99] using SWIR bands were proposed to obtain relative estimates of residue covers for multispectral and hyperspectral remote sensing data, respectively. Daughtry et al. evaluated several remote sensing indices to measure crop residue cover and utilized them to categorize soil tillage intensity in [98]. In [99], Daughtry and Hunt discussed the effects of water content on remote sensing-based residue estimation. Serbin et al. proposed a new index, the Shortwave Infrared Normalized Difference Residue Index (SINDRI) [100], which utilizes ASTER bands 6 and 7. SINDRI yielded more robust results than NDTI in wet conditions. In [97], Galloza

et al. investigated empirical models using NDTI and CAI of multiple remote sensing platforms, including Landsat TM, EO-1 ALI, Hyperion and airborne SpecTIR, and corresponding field transect data for reference. The airborne hyperspectral data which have a high signal to noise ratio (SNR) provided more capability for estimating the residue cover associated with in-situ measurements, but required extensive ground reference information, which is costly and logistically difficult to obtain over large regions.

The NASA Hyperion sensor on EO-1 has provided the foundation for upcoming spaceborne missions, as well as spatial coverage of the Earth′s surface that are limited in airborne hyperspectral sensors [102]. While the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), operated by NASA/JPL, is a state-of-the-art sensor among the airborne platforms and provides high SNR (500:1) data, Hyperion has low SNR (50:1), which is not adequate for a variety of scientific disciplines [103]. Low SNR and significant striping occurs in the SWIR bands, which contain useful absorption features for crop residue estimation [93,94]. Noise and stripes caused by inter-detector calibration differences in Hyperion data cannot be adequately removed by destriping strategies. Spectral unmixing, which finds extreme or pure pixels and then computes the corresponding fractional abundances, associates data with meaningful and noisy components [13]. Also, the abundances of the component can be combined with in-situ measurements to quantify the coverage fractions. In this chapter, spectral unmixing techniques are qualitatively and quantitatively compared to empirical models based on CAI to estimate crop residue with both airborne and spaceborne hyperspectral remotely sensed data. First, residue cover maps are created using CAI and linear/manifold learning-based spectral unmixing methods. Then, empirical models are developed using the estimated residue cover maps from image data and corresponding ground measured residue cover via field campaigns. Finally, issues of the index-based residue estimation model with recent noisy and heavily striped Hyperion data, as well as the advantages of manifold learning-based unmixing methods compared to the linear unmixing model are addressed and discussed.

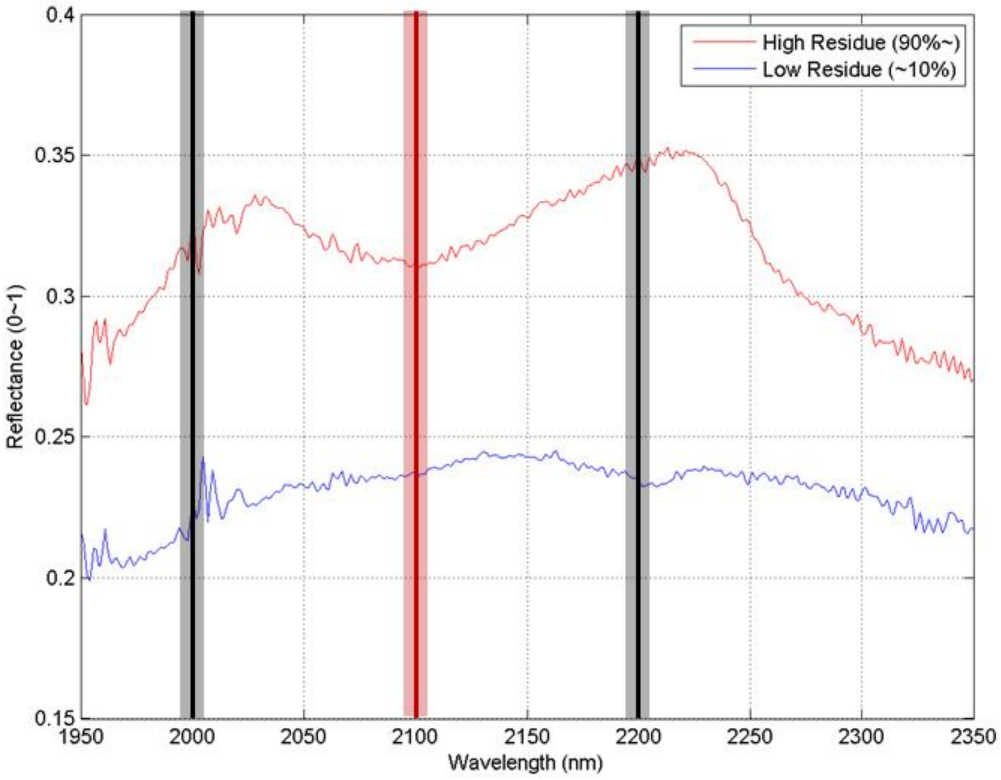


Figure 5.1. Absorption features of crop residue

5.2 Methodology

5.2.1 Cellulose Absorption Index (CAI)

Crop residue cover is linearly related to the depth of the broad absorption feature near 2100nm associated with cellulose as shown in Fig. 5.1. CAI, which is based on the spectral difference between adjacent bands near 2000nm and 2200nm, estimates the relative amount of crop residue as follows [94]:

$$\mathbf{CAI} = 0.5 \times (\mathbf{R}_{2.0} + \mathbf{R}_{2.2}) - \mathbf{R}_{2.1} \tag{5.1}$$

where $\mathbf{R}_{2.0}, \mathbf{R}_{2.1}, \mathbf{R}_{2.2}$ are the reflectance values in bands centered at 2000, 2100, and 2200nm, respectively. The adjacent bands around these wavelengths are averaged to improve stability.

5.2.2 Spectral Unmixing for Residue Estimation

In remotely sensed data, most pixels are highly mixed with several distinct materials for reasons such as resolution trade-off, intimate mixtures, multiple interferences, etc [13–16]. Spectral unmixing assumes that the surface is dominated by a small number of these substances and can be modeled by representing them at sub-pixel levels [13]. For these reasons, spectral unmixing provides the capability to quantitatively estimate residue cover as an alternative approach to the CAI method.

In this study, N-FINDR and fully constrained linear unmixing are used for baseline endmember extraction and abundance estimation, respectively. N-FINDR is one of the most popular, automatic EE algorithms (summarized in Chapter 3). PCA, the global manifold method ISOMAP, and the local manifold method LLE are compared to reduce the dimensional space for N-FINDR. Fully constrained linear unmixing is efficient for estimating physically meaningful fractional abundances as described in Chapter 1.

5.3 Data Acquisition for Hyperspectral Mapping of Residue Cover

5.3.1 Field Reference Data: Purdue University

This study was conducted over an agricultural area (ACRE) near Purdue University in Indiana, USA, where corn and soybeans are the dominant crops. Field campaigns were performed to collect residue cover with a hand-held GPS unit in selected fields during the fall of 2008 and 2012.

In 2008, the traditional field line-point transect, a standard technique for measuring the crop residue cover used by the United States Department of Agriculture

(a)

(b)

(c)

Figure 5.2. (a) line-point method and (b), (c) photographic acquisition technique using a digital camera mounted on a boom

(USDA) Natural Resources Conservation Service (NRCS) [104], was used at 38 locations. The line-point method uses systems of cross-hairs, grid points, or a dot matrix to define points where the presence or absence of residue is determined as shown in Fig. 5.2(a) [104].

The photographic technique using a digital camera mounted on a boom was adapted to determine the residue coverage at 27 locations in 2012. Nadir looking photos were acquired at approximately 7m height (see Fig. 5.2(b)). The pictures (e.g. Fig. 5.2(c)) were processed via SamplePoint software, which enables the measurement of cover from nadir looking images by superimposing a systematic or random array [105], to determine the proportion of crop residue cover in the field.

5.3.2 Remote Sensing Data for Residue Cover Evaluation Experiments

Hyperspectral data were collected by the airborne sensor SpecTIR and spaceborne sensor Hyperion on EO-1 over ACRE on November 3, 2008 and November 17, 2012, respectively (see Fig. 5.3). There were no heavy rain events before the remote sensing data acquisitions. Pure crop residue is more prevalent immediately following the harvest prior to significant tillage activity, whereas pure soil is easier to obtain near the end of the tillage season once farmers have prepared the spring seeding [96].

The SpecTIR data were recorded at 2m spatial resolution and 178 spectral bands of 10nm width, ranging from 400 to 2450nm. The Hyperion data were comprised of 196 bands of the original 224 bands after removing the water absorption bands in the vicinity of 1400 and 1900nm and the bands with low SNR. Unfortunately, the striping in Hyperion data has increased as the sensor has aged, and is particularly strong in the SWIR bands. The stripes removed using a local window normalization approach had minimal impact on the spectral response. Both datasets were atmospherically corrected to convert the radiance to reflectance using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) radiative transfer code in ENVI and then geometrically corrected. The spatial resolution of the airborne data was

(a) (b)

Figure 5.3. Color composite images of study area over ACRE collected by (a) SpecTIR (R/G/B: 650/555/445nm, Fall 2008) and (b) Hyperion (R/G/B: 650/559/447nm, Fall 2012) (Images acquired over different areas)

degraded to 30m by bilinear interpolation to observe the mixing phenomena, and to match the spatial resolution of Hyperion.

5.4 Experimental Unmixing Results

5.4.1 Airborne Hyperspectral Data (Fall 2008) for Unmixing Experiments

A. Qualitative evaluation of unmixing results

The SpecTIR data contain 10 potential endmembers estimated by the HySime algorithm [79], which is eigen decomposition-based and fully automatic, from the airborne hyperspectral data collected by SpecTIR. The SpecTIR data have high SNR (500:1) and are not as affected by atmospheric attenuation as the spaceborne Hyperon data. Fig. 5.4(a) shows a residue cover map obtained by CAI, and Figs. 5.4(b)-5.4(d) illustrate fractional abundance maps of the residue component by PCA, ISOMAP, and LLE-based unmixing methods, respectively. The brighter pixels are associated with a higher residue cover. Unlike CAI (Fig. 5.4(a)), unmixing results (Fig. 5.4(b)-5.4(d)), especially in the NW fields, indicated a within field variability that might be associated with other materials such as green vegetation, wet areas, etc. Some mid-coverage fields obtained using LLE for feature reduction over-estimated the residue cover compared to those of the ISOMAP unmixing method. While unmixing extracts these materials as individual components, CAI focuses only on band combinations associated with the cellulose absorption feature, and other contributing factors cannot be discerned directly. Manifold learning-based unmixing results (Fig. 5.4(c), 5.4(d)) also exhibited a greater within field variability and captured more detailed characteristics in some fields than did PCA-based unmixing (Fig. 5.4(b)).

B. Quantitative evaluation of residue estimates

As a quantitative validation, linear regression models based on in-situ measurements and corresponding image derived values of residue cover were developed as shown in Fig. 5.5. The CAI values were scaled between 0 and 1 to be consistent with fully constrained unmixing. All the regression models had high R^2 values. In general, the results from the various models disagreed more in fields with high residue

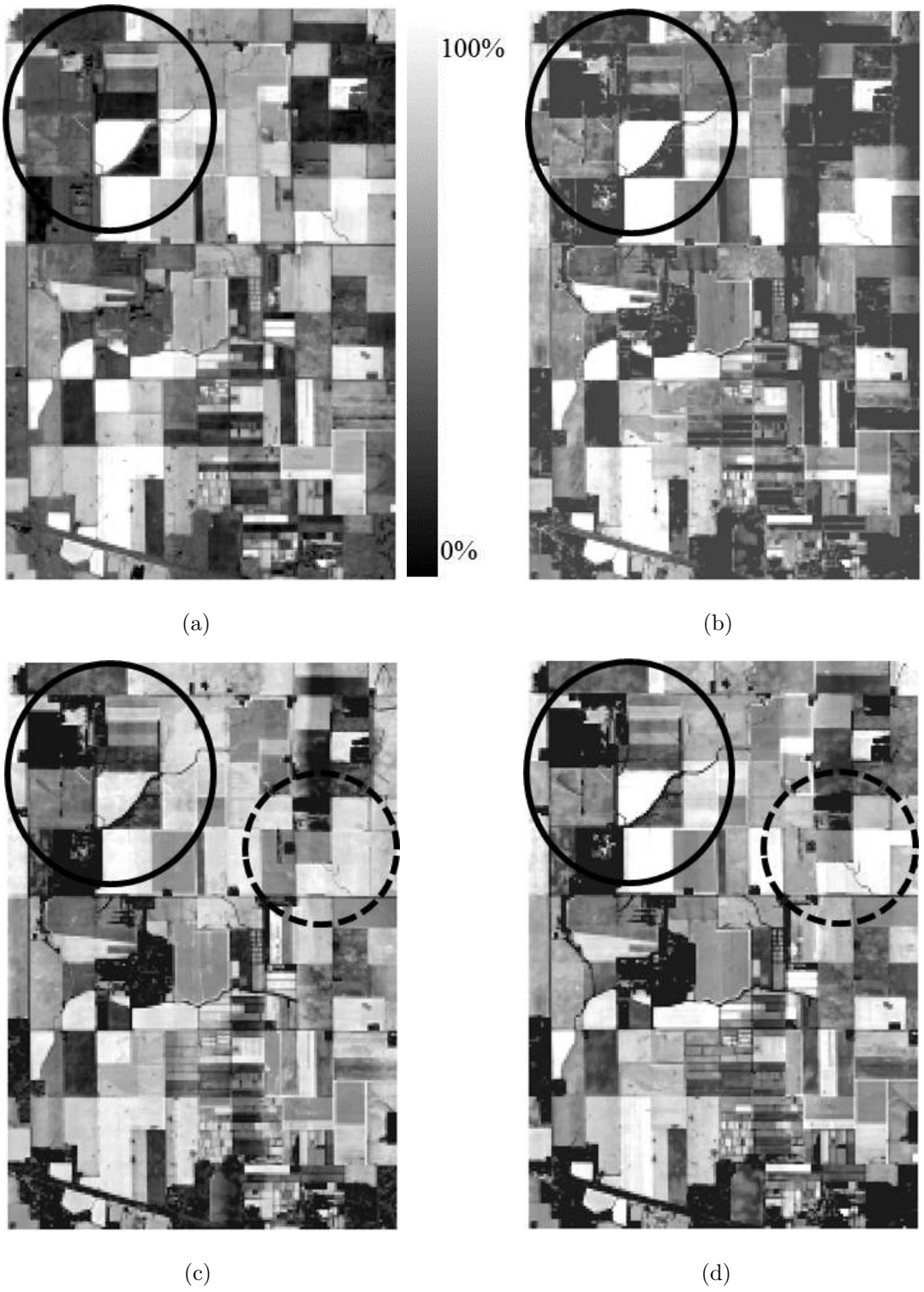
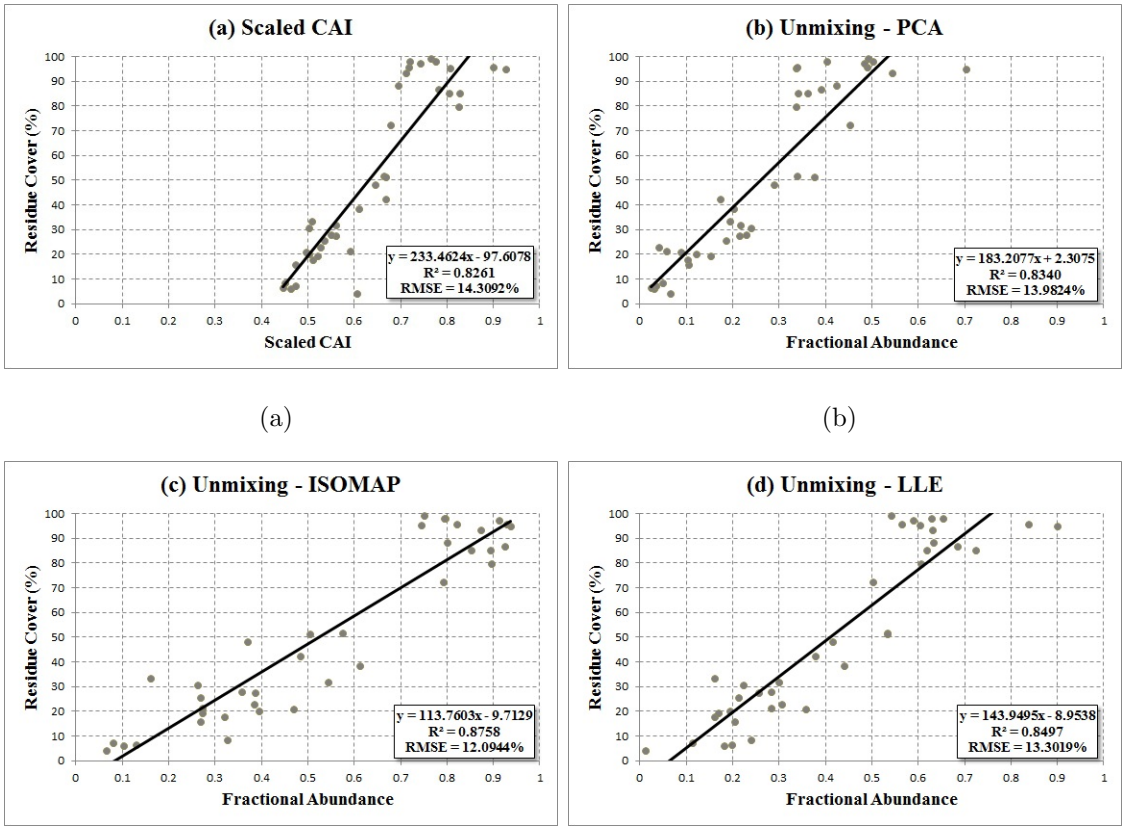


Figure 5.4. Estimated residue cover maps of SpecTIR data by (a) CAI, (b) Unmixing (PCA), (c) Unmixing (ISOMAP) and (d) Unmixing (LLE)

coverage than in fields with low residue coverage. High residue cover is due to corn stover, which has greater variability in reflectance, in part due to variation of residue water content. All the unmixing-based models (Fig. 5.5(b)-5.5(d)) had higher R^2 and lower RMSE values than the CAI-based model. The models derived from unmixing covered the full range of the ground based measurements, and the slope of the models for ISOMAP and LLE was approximately one (note: the slope in the scatter plots was 100 due to the different units between fraction and coverage). As shown in Fig. 5.5(c),5.5(d), the manifold learning-based unmixing results also had higher R^2 and lower RMSE values than the PCA-based model results illustrated in Fig. 5.5(b). Fig. 5.6 shows RMSE over the range of residue cover values between in-situ and estimated residue cover resulting from the regression models. While CAI generally has high accuracy in the low-coverage fields, unmixing methods have relatively low RMSE as the residue cover increases. For unmixing techniques, standard deviations of the residuals are also lower than for CAI based models, which indicates that the results are robust for residue coverage.

Finally, it should be noted that while CAI can be interpreted relative to the absorption feature of cellulose in the electromagnetic spectrum, it must be calibrated with ground measurements. Unmixing can provide estimates of crop residue coverage without the corresponding ground data due to the nature of the endmembers and the constraints of the linear mixing model. Since a fully constrained linear unmixing was used, the abundances had a range of 0 to 1 (e.g., 100% (pure) residue cover should be close to 1 and 0% residue cover (pure soil) should be close to 0). In results where PCA is used for dimensionality reduction (Fig. 5.5(b)), fractional abundances generally have low pixel values (less than 0.5) compared to those of manifold learning based methods (Fig. 5.5(c),5.5(d)). This indicates that manifold learning-based methods provided better characterization of the data in a low dimensional space and selected a more accurate residue endmember than could PCA.



(a) (b)

(c) (d)

Figure 5.5. Linear regression models between ground measures and (a) scaled CAI, (b) Unmixing (PCA), (c) Unmixing (ISOMAP), (d) Unmixing (LLE) from SpecTIR data

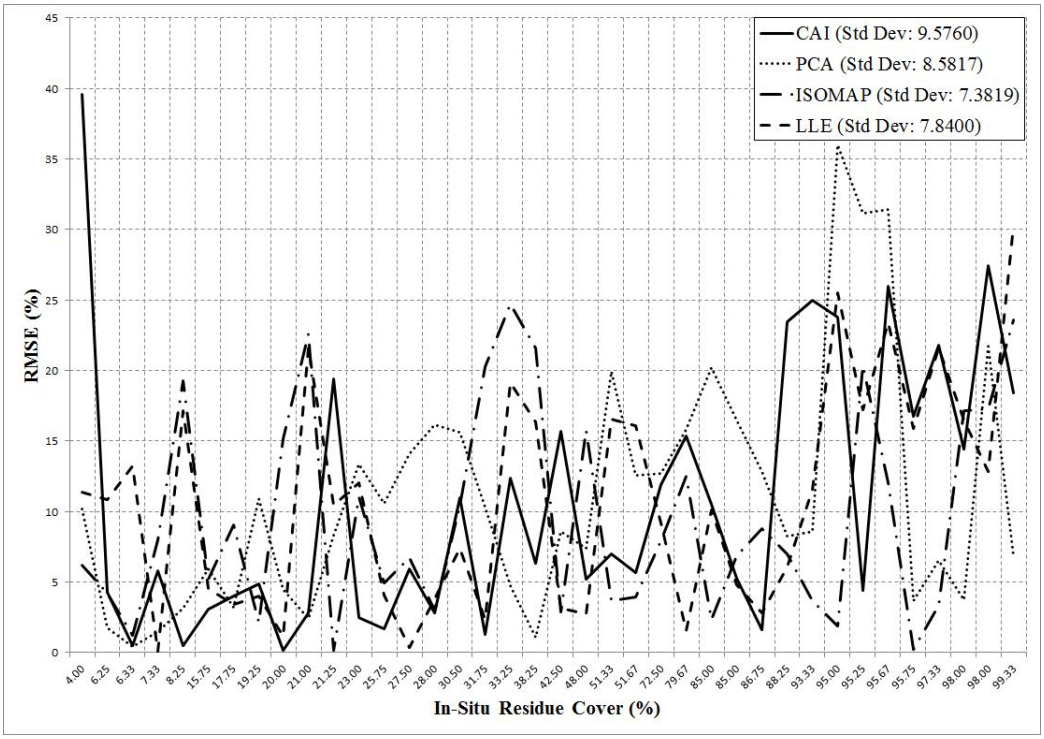


Figure 5.6. RMSE comparison between in-situ and estimated residue cover of the selected fields

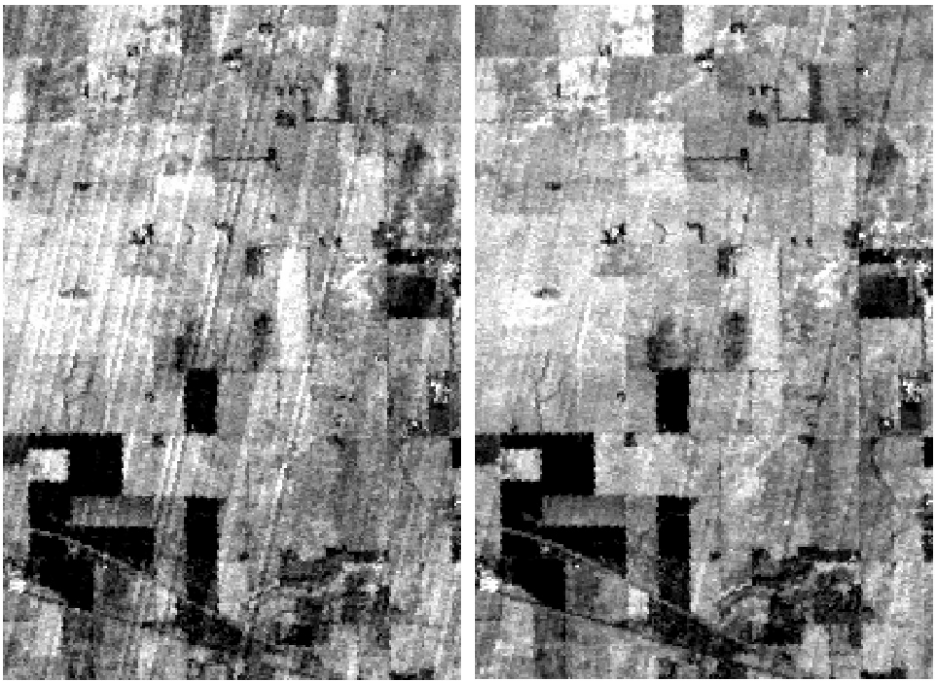
5.4.2 Hyperion Data (Fall 2012) for Unmixing Experiments

A. Qualitative evaluation of unmixing results

Although airborne hyperspectral data have high SNR and are not impacted by atmospheric attenuation, the acquisitions are limited and costly. Over the past decade, users have had opportunities to access data acquired by Hyperion, in preparation for the launch of missions such as the Japanese HISUI, European Space Agency EnMAP, and NASA HypsIRI. Unfortunately, recent Hyperion imagery has contained more stripes than the historical Hyperion data, particularly in the SWIR. $\mathbf{R}_{2.0}, \mathbf{R}_{2.1}, \mathbf{R}_{2.2}$ of Hyperion that are used for CAI visually appear to have been adequately corrected (see Fig. 5.7(a)5.7(c)), but the resulting CAI image contains many artifacts (see Fig. 5.7(d)). The 2000nm band was not utilized in the estimate due to noise. How-

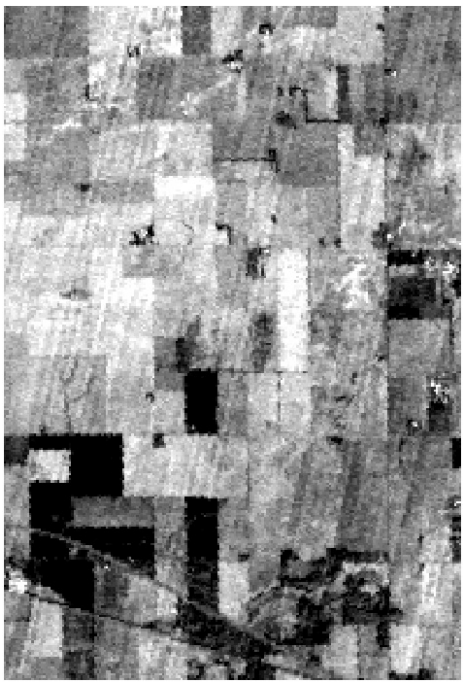
ever, meaningful information from the CAI-based image could still not be obtained since the stripes and the noise associated with $\mathbf{R}_{2.0}, \mathbf{R}_{2.1}, \mathbf{R}_{2.2}$ were exacerbated.

The Hyperion dataset also contains 10 endmembers. Fig. 5.8(a) represents a fractional abundance map of the residue component by the PCA-based unmixing method. Stripes were still present, but the resulting image was visually better than the CAI image (Fig. 5.7(d)) for interpreting residue cover. As shown in Fig. 5.8(d), the stripes and the noise components were extracted as a separate component and then used to compute the fractional abundances in a linear mixing model. Since the stripes and the noise generally have extreme pixel values, and are present throughout the scene, they are good endmember candidates due to property of the endmembers.

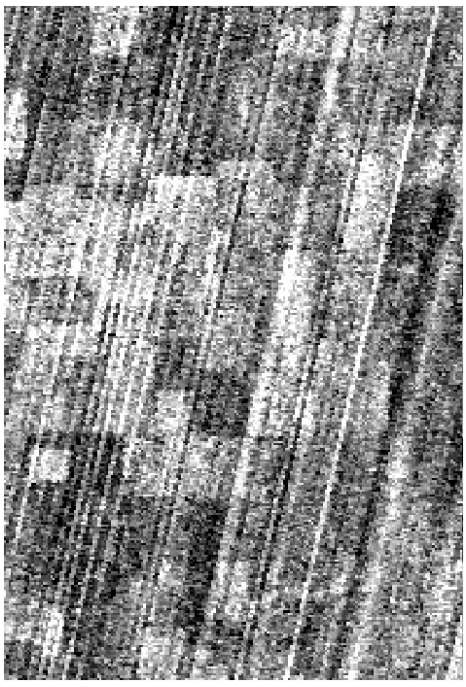


(a)

(b)

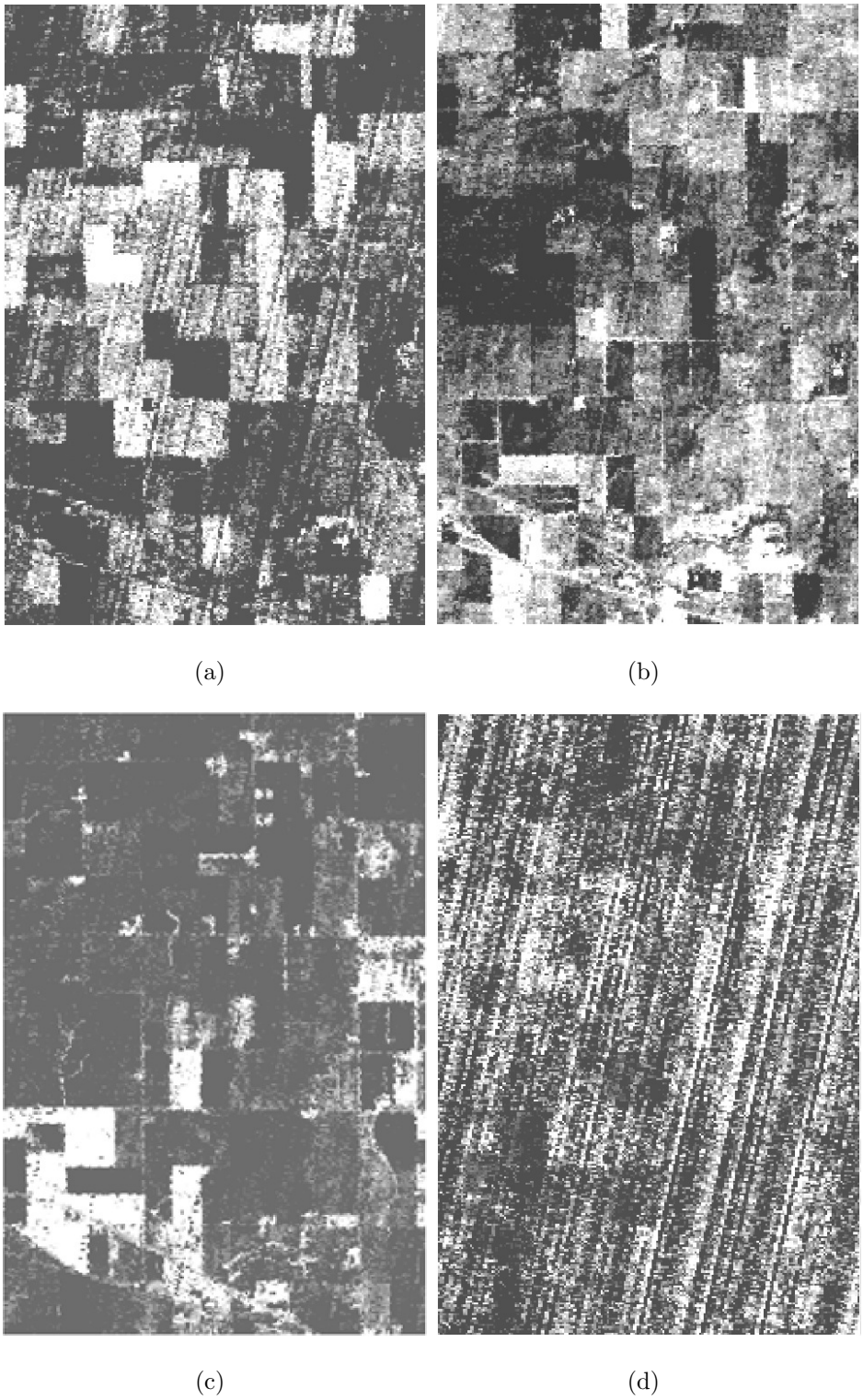


(c)



(d)

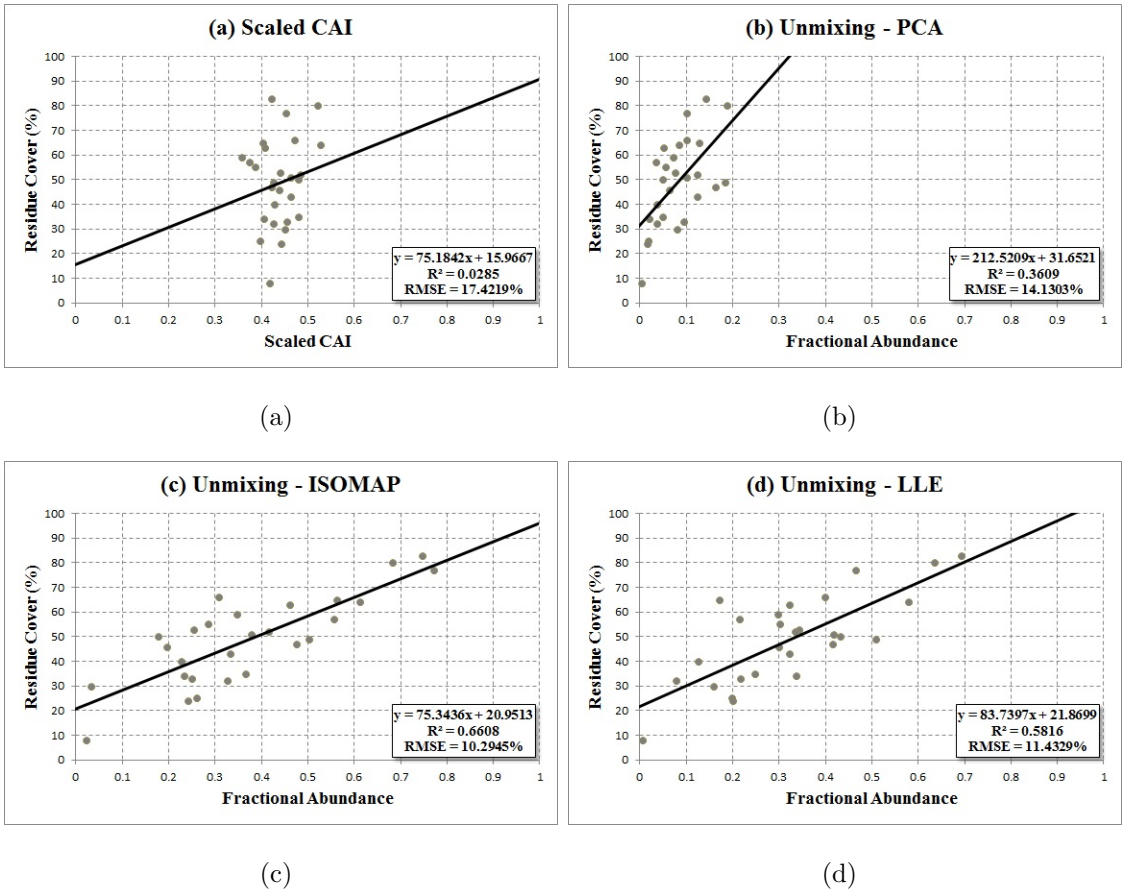
Figure 5.7. (a) 1990nm, (b) 2100nm, (c) 2200nm band of Hyperion, and (d) CAI image



(a) (b)

(c) (d)

Figure 5.8. Abundance maps of (a) crop residue and (b) soil, (c) vegetation, and (d) noise component extracted from Hyperion



(a) (b)

(c) (d)

Figure 5.9. Linear regression models between ground measures and (a) scaled CAI, (b) Unmixing (PCA), (c) Unmixing (ISOMAP), (d) Unmixing (LLE) from Hyperion data

B. Quantitative evaluation of residue estimates

Empirical models were developed in a similar way to those from airborne data analysis. Based on the visual inspection in Fig. 5.7(d), a regression model based on CAI and ground measurements would not be expected useful information (also note Fig. 5.9(a)). Both PCA and manifold learning-based unmixing methods (Fig. 5.9(b)-5.9(d)) had better R^2 and RMSE than the CAI-based regression model. Since airborne data have much better SNR and original spatial resolution than Hyperion, R^2 values of airborne models were generally much higher (> 0.8). The study area, field data acquisition method, and field conditions were also different between 2008 and 2012,

so models were not really comparable. However, the trends for the two models are similar to those obtained from airborne data. The CAI based-model has the narrowest dynamic range. The fractional abundances obtained via PCA feature extraction and unmixing method are biased and the range of estimated abundances is very small (less than 0.2) compared to the manifold learning methods, which ranged from 0 to 0.8. The slope of PCA which indicates the physical agreement is much higher than that of the manifold learning-based models. Manifold learning-based methods had slightly higher R^2 values, with a slope closer to one and lower RMSE than PCA in the SpecTIR dataset, but they significantly outperformed PCA as well as CAI in the noisy Hyperion data.

5.5 Conclusions

In this chapter, spectral unmixing techniques using hyperspectral remotely sensed data with high and low SNR were investigated to estimate crop residue in agricultural areas. In general, both CAI and unmixing models performed well if the image data did not contain stripes and had high SNR. However, the unmixing methods significantly outperformed the CAI method in recent Hyperion data since noise was embedded in the CAI values via band combinations. Further, manifold learning yielded a more representative residue endmember than PCA, the linear feature extraction method.

Three conclusions were derived from this case study.

- Spectral unmixing techniques can mitigate the impact of noise and stripes, particularly in noisy data collected by pushbroom sensors since the noise and the stripes associated with generally extreme spectral signatures were extracted as individual components. If a pure spectral signature of crop residue was selected, it resulted in an improved visual qualitative map and a quantitatively more accurate residue map, based on the field data collected in this study.

- Although manifold learning is computationally expensive, it is more efficient for finding a better residue endmember and for estimating residue cover than is

PCA. The high computational overhead can be mitigated by landmark-based manifold learning approaches as discussed in the previous chapters.

- Fully constrained unmixing provides physically more meaningful results compared to CAI. Incorporating ground data is necessary for model development, but fully constrained fractional abundances themselves are sufficiently interpretable.

Spectral unmixing represents the data in terms of continuous values. It can exploit the physical proportions of interesting targets in highly mixed remote sensing data if the proper endmembers are selected directly from the image, field or spectral library, which are straightforward to develop spectral characteristics of the target. However, to develop appropriate remote sensing indices such as CAI, the distinguishing spectral characteristics (e.g., absorption features) of the target must be adequately investigated. This also requires a strong background in spectroscopy.

The water content of soil and crop residue impacts remotely sensed estimates of crop residue cover [99, 100]. In this study, since there were no heavy rain events reported prior to the data acquisitions, it was unable to determine whether the unmixing models were robust for the residue water content and soil moisture. Also, it should be noted that the unmixing techniques had relatively higher errors in low residue cover fields than the CAI model (Fig. 5.6) since only fractional abundances of pure residue endmember were involved in developing the empirical models. If the fractional abundances of the residue can be combined with a soil component (Fig. 5.8(b)) in the empirical models, the model accuracy in the low residue fields might be increased. It should be noted that fields with low residue cover are more critical to soil and nutrient loss from rainfall, as shown in [106]. Further research over a broader range of conditions is needed to improve the residue predictions, possibly jointly exploiting both the CAI and unmixing approaches based on the hyperspectral data.

6. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

6.1 Summary

Manifold learning-based spectral unmixing of hyperspectral remote sensing data was investigated in this dissertation. The aim was to propose an intelligent landmark selection method that could deal with high computational overhead of manifold learning and to develop more representative manifolds by an active landmark sampling strategy. The primary contributions of the dissertation are summarized for each chapter as follows:

- **Chapter 3:** A new landmark selection method and hybrid manifold approach were investigated. The resulting landmark points identified by the proposed LH method were acquired at lower computational cost and had spectrally and spatially low redundancy compared to in the traditional methods since the proposed landmark selection method exploited both the spectral and spatial relationships between adjacent pixels. The proposed hybrid manifold could further reduce the computational burden by being integrated with the LH method. Globally selected backbone points and local approximations could preserve both the global and local characteristics of the geometry.

  [Related publications]

  [73] J. Chi and M. M. Crawford. Landmark selection using homogeneity on nonlinear manifolds for unmixing hyperspectral data. *Proc. 2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Munich, Germany, July 22-27, pp. 1373-1376, 2012.

[74] J. Chi and M. M. Crawford. Selection of landmark points on nonlinear manifolds for spectral unmixing using local homogeneity. *IEEE Geoscience and Remote Sensing Letters*, 10(4): 711-715, 2013.

- **Chapter 4:** Active learning heuristics were incorporated into the landmark selection process. Additional landmarks identified by active sampling strategy were selected to increase the manifold stability and representation for spectral unmixing. A sampling method based on spatially nonoverlapping, sequentially subdivided subsets was advantageous for efficiently selecting the additional landmarks. The proposed sampling method was successfully incorporated into the active sampling framework and developed both spectrally and spatially high diverse landmark points with fewer learning steps.

  [Related publications]

  [80] J. Chi and M. M. Crawford. Active landmark sampling for manifold learning based spectral unmixing. *IEEE Geoscience and Remote Sensing Letters*, 2013, In revision.

- **Chapter 5:** The spectral unmixing technique was investigated to estimate crop residue cover. For the airborne data, both CAI and unmixing-based methods produced reasonable results with high agreement. In a noisy image, however, the unmixing technique, especially the manifold learning-based approaches, developed more accurate empirical models than the CAI-based model due to the impact of extreme values, and yielded visually better residue coverage maps. Additionally, fractional abundances are sufficiently interpretable without incorporating in-situ measurements.

  [Related publications]

  [92] J. Chi and M. M. Crawford. Spectral unmixing based crop residue estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2013, In revision.

6.2 Future work

Although this dissertation addressed several issues inherent in existing landmark sampling methods and unmixing techniques, remaining challenges motivate extensions of this research:

1. Manifold learning-based spectral unmixing captures some forms of nonlinearity of the data in the feature extraction stage, and endmembers are extracted from the manifold space. Although the fully constrained linear unmixing to compute the fractional abundances can estimate physically meaningful proportions of the extracted endmembers, limitations remain when attempting to solve problems where the linear model is not adequate. In the machine learning community, kernel-based algorithms in Hilbert space have gained attention and achieved notable success in addressing nonlinear problems in various applications. Kernel-based approaches may provide a capability to resolve the issue of nonlinear mixture analysis while maintaining the constraints of the fully constrained linear mixing model.

2. Since active landmark sampling is a relatively new concept which was motivated by active learning in supervised classification, further research should be conducted to address both the classification and unmixing tasks. More advanced sampling methods to better explore the spectral and spatial context are worthwhile topics.

3. In residue research, the soil endmember component is highly related to the crop residue endmember. Since the soil endmember is more likely to be associated with low residue covered fields, a combination of both soil and residue endmembers might be able to increase the agreement of the regression model. Addressing soil moisture as an influence on the residue prediction model remains a challenging problem due to the local variability of soil moisture and residue water content.

LIST OF REFERENCES

LIST OF REFERENCES

[1] R. N. Clark. Spectroscopy of rocks and minerals, and principles of spectroscopy. *Manual of Remote Sensing*, 3:3–58, 1999.

[2] M. E. Martin and J. D. Aber. High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes. *Ecological Applications*, 7(2):431–443, 1997.

[3] J. F. Mustard and J. M. Sunshine. Spectral analysis for earth science: investigations using remote sensing data. *Remote Sensing for the Earth Sciences: Manual of Remote Sensing*, 3:251–307, 1999.

[4] J. Solomon and B. Rock. Imaging spectrometry for earth remote sensing. *Science*, 228:1147–1153, 1985.

[5] G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1):55–63, 1968.

[6] D. Landgrebe. Hyperspectral image data analysis. *IEEE Signal Processing Magazine*, 19(1):17–28, 2002.

[7] L. O. Jimenez and D. A. Landgrebe. Hyperspectral data analysis and supervised feature reduction via projection pursuit. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6):2653–2667, 1999.

[8] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina. Improved manifold coordinate representations of large-scale hyperspectral scenes. *IEEE Transactions on Geoscience and Remote Sensing*, 44(10):2786–2803, 2006.

[9] S. R. Sandmeier, E. M. Middleton, D. W. Deering, and W. Qin. The potential of hyperspectral bidirectional reflectance distribution function data for grass canopy characterization. *Journal of Geophysical Research*, 104(D8):9547–9560, 1999.

[10] D. G. Goodin, J. Gao, and G. M. Henebry. The effect of solar illumination angle and sensor view angle on observed patterns of spatial structure in tallgrass prairie. *IEEE Transactions on Geoscience and Remote Sensing*, 42(1):154–165, 2004.

[11] J. F. Mustard, L. Li, and G. He. Nonlinear spectral mixture modeling of lunar multispectral data: Implications for lateral transport. *Journal of Geophysical Research: Planets (1991–2012)*, 103(E8):19419–19425, 1998.

[12] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina. Improvements to land-cover and invasive species mapping from hyperspectral imagery in the virginia coast reserve. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 6, pages 4180–4183, Anchorage, Alaska, 2004. IEEE.

[13] N. Keshava and J. F. Mustard. Spectral unmixing. *IEEE Signal Processing Magazine*, 19(1):44–57, 2002.

[14] K. Segl, L. Guanter, H. Kaufmann, J. Schubert, S. Kaiser, B. Sang, and S. Hofer. Simulation of spatial sensor characteristics in the context of the enmap hyperspectral mission. *IEEE Transactions on Geoscience and Remote Sensing*, 48(7):3046–3054, 2010.

[15] C. Ekaterina and C. Véronique. Spectral un-mixing of natural surfaces scenarios. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 4491–4494, Honolulu, Hawaii, 2010. IEEE.

[16] J. B. Adams, M. O. Smith, and P. E. Johnson. Spectral mixture modeling: A new analysis of rock and soil types at the viking lander 1 site. *Journal of Geophysical Research*, 91(B8):8098–8112, 1986.

[17] A. Plaza, P. Martínez, R. Perez, and J. Plaza. A comparative analysis of endmember extraction algorithms using aviris hyperspectral imagery. In *Summaries of the 11th JPL Airborne Earth Science Workshop*, 2002.

[18] J. Plaza, A. Plaza, P. Martínez, and R. Pérez. H-comp: A tool for quantitative and comparative analysis of endmember identification algorithms. In *2003 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 1, pages 291–293, Toulouse, France, 2003. IEEE.

[19] A. Plaza, P. Martinez, R. Pérez, and J. Plaza. A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 42(3):650–663, 2004.

[20] J. C. Harsanyi, W. H. Farrand, and C.-I. Chang. Detection of subpixel signatures in hyperspectral image sequences. In *Proceedings of the American Society for Photogrammetry and Remote Sensing*, pages 236–247, 1994.

[21] W. H. Farrand, R. B. Singer, and E. Merényi. Retrieval of apparent surface reflectance from aviris data: A comparison of empirical line, radiative transfer, and spectral mixture methods. *Remote Sensing of Environment*, 47(3):311–321, 1994.

[22] J. W. Boardman, F. A. Kruse, and R. O. Green. Mapping target signatures via partial unmixing of aviris data. *5th JPL Airborne Earth Science Workshop*, 1:23–26, 1995.

[23] C.-I. Chang. *Hyperspectral imaging: Techniques for spectral detection and classification*. Springer, 2003.

[24] D. C. Heinz. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(3):529–545, 2001.

[25] C. C. Borel and S. Gerstl. Nonlinear spectral mixing models for vegetative and soil surfaces. *Remote Sensing of Environment*, 47(3):403–416, 1994.

[26] S. G. Herzog and J. F. Mustard. Reflectance spectra of five-component mineral mixtures: Implications for mixture modeling. In *Lunar and Planetary Institute Science Conference Abstracts*, volume 27, page 535, 1996.

[27] N. Keshava. A survey of spectral unmixing algorithms. *Lincoln Laboratory Journal*, 14(1):55–78, 2003.

[28] W. Liu and E. Y. Wu. Comparison of non-linear mixture models: Sub-pixel classification. *Remote Sensing of Environment*, 94(2):145–154, 2005.

[29] J. Broadwater and A. Banerjee. A comparison of kernel functions for intimate mixture models. In *First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHISPERS'09*, pages 1–4. IEEE, 2009.

[30] J. Plaza, A. Plaza, R. Perez, and P. Martinez. On the use of small training sets for neural network-based characterization of mixed pixels in remotely sensed hyperspectral images. *Pattern Recognition*, 42(11):3032–3045, 2009.

[31] Y. Altmann, A. Halimi, N. Dobigeon, and J. Tourneret. Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery. *IEEE Transactions on Image Processing*, 21(6):3017–3025, 2012.

[32] J. He, L. Zhang, Q. Wang, and Z. Li. Using diffusion geometric coordinates for hyperspectral imagery representation. *IEEE Geoscience and Remote Sensing Letters*, 6(4):767–771, 2009.

[33] F. Mathieu, C. Jocelyn, and B. Jón Atli. Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas. *EURASIP Journal on Advances in Signal Processing*, 11, 2009.

[34] M. M. Crawford, L. Ma, and W. Kim. Exploring nonlinear manifold learning for classification of hyperspectral data. In *Optical Remote Sensing*, pages 207–234. Springer, 2011.

[35] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce. Locality-preserving discriminant analysis in kernel-induced feature spaces for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 8(5):894–898, 2011.

[36] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce. Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 50(4):1185–1198, 2012.

[37] R. Heylen, D. Burazerovic, and P. Scheunders. Non-linear spectral unmixing by geodesic simplex volume maximization. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):534–542, 2011.

[38] R. Heylen and P. Scheunders. Calculation of geodesic distances in nonlinear mixing models: Application to the generalized bilinear model. *IEEE Geoscience and Remote Sensing Letters*, 9(4):644–648, 2012.

[39] N. Hoang, C. Richard, P. Honeine, and C. Theys. Hyperspectral image unmixing using manifold learning methods derivations and comparative tests. In *2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3086–3089, Munich,Germany, 2012. IEEE.

[40] A. A. Green, M. Berman, P. Switzer, and M. D. Craig. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1):65–74, 1988.

[41] A. Ifarraguerri and C.-I. Chang. Multispectral and hyperspectral image analysis with convex cones. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2):756–770, 1999.

[42] M. E. Winter. N-findr: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*, pages 266–275. International Society for Optics and Photonics, 1999.

[43] A. Plaza and C.-I. Chang. An improved n-findr algorithm in implementation. In *Defense and Security*, pages 298–306. International Society for Optics and Photonics, 2005.

[44] C.-I. Chang, C.-C. Wu, W. Liu, and Y.-C. Ouyang. A new growing method for simplex-based endmember extraction algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 44(10):2804–2819, 2006.

[45] Q. Du, N. Raksuntorn, N. H. Younan, and R. L. King. Variants of n-findr algorithm for endmember extraction. In *Remote Sensing*, pages 71090G–71090G. International Society for Optics and Photonics, 2008.

[46] M. Zortea and A. Plaza. A quantitative and comparative analysis of different implementations of n-findr: A fast endmember extraction algorithm. *IEEE Geoscience and Remote Sensing Letters*, 6(4):787–791, 2009.

[47] C.-I. Chang, C.-C. Wu, and C.-T. Tsai. Random n-finder (n-findr) endmember extraction algorithms for hyperspectral imagery. *IEEE Transactions on Image Processing*, 20(3):641–656, 2011.

[48] K. Staenz, T. Szeredi, and J. Schwarz. Isdas: A system for processing/analyzing hyperspectral data. *Canadian Journal of Remote Sensing*, 24(2):99–113, 1998.

[49] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington. ICE: A statistical approach to identifying endmembers in hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 42(10):2085–2095, 2004.

[50] A. Singh and A. Harrison. Standardized principal components. *International Journal of Remote Sensing*, 6(6):883–896, 1985.

[51] L. Eklundh and A. Singh. A comparative analysis of standardised and unstandardised principal components analysis in remote sensing. *International Journal of Remote Sensing*, 14(7):1359–1370, 1993.

[52] C.-I. Chang. *Hyperspectral Data Processing: Algorithm Design and Analysis*. Wiley-Interscience, 2013.

[53] A. Plaza, P. Martinez, R. Pérez, and J. Plaza. Spatial/spectral endmember extraction by multidimensional morphological operations. *IEEE Transactions on Geoscience and Remote Sensing*, 40(9):2025–2041, 2002.

[54] D. M. Rogge, B. Rivard, J. Zhang, A. Sanchez, J. Harris, and J. Feng. Integration of spatial-spectral information for the improved extraction of endmembers. *Remote Sensing of Environment*, 110(3):287–303, 2007.

[55] S. Mei, M. He, Z. Wang, and D. Feng. Spatial purity based endmember extraction for spectral mixture analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 48(9):3434–3445, 2010.

[56] C. Lee and D. A. Landgrebe. Feature extraction based on decision boundaries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):388–400, 1993.

[57] D. A. Landgrebe. *Signal theory methods in multispectral remote sensing*. John Wiley and Sons, Inc., 2003.

[58] M. Pal and G. M. Foody. Feature selection for classification of hyperspectral data by svm. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5):2297–2307, 2010.

[59] K. Fukunaga. *Introduction to statistical pattern recognition*. Access Online via Elsevier, 1990.

[60] A. Cheriyadat and L. M. Bruce. Why principal component analysis is not an appropriate feature extraction method for hyperspectral data. In *2003 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 6, pages 3420–3422, Toulouse, France, 2003. IEEE.

[61] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina. Exploiting manifold geometry in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):441–454, 2005.

[62] C.-M. Chen, G. F. Hepner, and R. R. Forster. Fusion of hyperspectral and radar data using the ihs transformation to enhance urban surface features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(1):19–30, 2003.

[63] J. Wang and C.-I. Chang. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6):1586–1600, 2006.

[64] J. B. Tenenbaum, V. De Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[65] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[66] L. K. Saul and S. T. Roweis. Think globally, fit locally: unsupervised learning of low dimensional manifolds. *The Journal of Machine Learning Research*, 4:119–155, 2003.

[67] F. Mathieu, C. Jocelyn, and B. Jón Atli. Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas. *EURASIP Journal on Advances in Signal Processing*, 2009, 2009.

[68] L. Ma, M. M. Crawford, and J. Tian. Local manifold learning-based k-nearest-neighbor for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4099–4109, 2010.

[69] Y. Chen, M. M. Crawford, and J. Ghosh. Improved nonlinear manifold learning for land cover classification via intelligent landmark selection. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 545–548. IEEE, 2006.

[70] V. De Silva and J. B. Tenenbaum. Sparse multidimensional scaling using landmark points. Technical report, Technical report, Stanford University, 2004.

[71] J. Silva, J. Marques, and J. Lemos. Selecting landmark points for sparse manifold learning. In *Advances in Neural Information Processing Systems*, pages 1241–1248, 2005.

[72] Trevor. H., Robert. T., and J. H. Friedman. *The elements of statistical learning*, volume 1. Springer New York, 2001.

[73] J. Chi and M. M. Crawford. Landmark selection using homogeneity on nonlinear manifolds for unmixing hyperspectral data. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1373–1376, Munich, Germany, 2012. IEEE.

[74] J. Chi and M. M. Crawford. Selection of landmark points on nonlinear manifolds for spectral unmixing using local homogeneity. *IEEE Geoscience and Remote Sensing Letters*, 10:711–715, 2013.

[75] N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[76] G. Chen and S.-E. Qian. Dimensionality reduction of hyperspectral imagery using improved locally linear embedding. *Journal of Applied Remote Sensing*, 1(1), 2007.

[77] R. N. Clark, G. A. Swayze, K. E. Livo, R. F. Kokaly, S. J. Sutley, J. B. Dalton, R. R. McDougal, and C. A. Gent. Imaging spectroscopy: Earth and planetary remote sensing with the usgs tetracorder and expert systems. *Journal of Geophysical Research*, 108(E12):5131, 2003.

[78] R. N. Clark and G. A. Swayze. Mapping minerals, amorphous materials, environmental materials, vegetation, water, ice and snow, and other materials: the usgs tricorder algorithm. In *Summaries of the Fifth Annual JPL Airborne Earth Science Workshop*, volume 1, pages 39–40. JPL Publication, 1995.

[79] J. M. Bioucas-Dias and J. M. P. Nascimento. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8):2435–2445, 2008.

[80] J. Chi and M. M. Crawford. Active landmark sampling for manifold learning based spectral unmixing. *IEEE Geoscience and Remote Sensing Letters (In Review)*, 2013.

[81] S. Dasgupta and D. Hsu. Hierarchical sampling for active learning. In *Proceedings of the 25th international conference on Machine learning*, pages 208–215. ACM, 2008.

[82] F. Olsson. A literature survey of active machine learning in the context of natural language processing. 2009.

[83] B. Settles. Active learning literature survey. *University of Wisconsin, Madison*, 2010.

[84] G. M. Foody, A. Mathur, C. Sanchez-Hernandez, and D. S. Boyd. Training set size requirements for the classification of a specific class. *Remote Sensing of Environment*, 104(1):1–14, 2006.

[85] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari. A survey of active learning algorithms for supervised remote sensing image classification. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):606–617, 2011.

[86] G. Healey and D. Slater. Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6):2706–2717, 1999.

[87] B. Thai and G. Healey. Invariant subpixel target identification in hyperspectral imagery. In *AeroSense'99*, pages 14–24. International Society for Optics and Photonics, 1999.

[88] W. Boardman. Inversion of imaging spectrometry data using singular value decomposition. In *1989 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 4, pages 2069–2072. IEEE, 1989.

[89] S. Danaher and E. O'mongain. Singular value decomposition in multispectral radiometry. *International Journal of Remote Sensing*, 13(9):1771–1777, 1992.

[90] M. Berger. *Geometry I and II*, volume 1. Springer, 1987.

[91] Q. Du, N. Raksuntorn, N. H. Younan, and R. L. King. End-member extraction for hyperspectral image analysis. *Applied Optics*, 47(28):F77–F84, 2008.

[92] J. Chi and M. M. Crawford. Spectral unmixing based crop residue estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (In Review)*, 2013.

[93] C. S. T. Daughtry. Discriminating crop residues from soil by shortwave infrared reflectance. *Agronomy Journal*, 93(1):125–131, 2001.

[94] C. S. T. Daughtry, E. R. Hunt Jr, and J. E. McMurtrey III. Assessing crop residue cover using shortwave infrared reflectance. *Remote Sensing of Environment*, 90(1):126–134, 2004.

[95] A. Bannari, A. Pacheco, K. Staenz, H. McNairn, and K. Omari. Estimating and mapping crop residues cover on agricultural lands using hyperspectral and IKONOS data. *Remote Sensing of Environment*, 104(4):447–459, 2006.

[96] A. Pacheco and H. McNairn. Evaluating multispectral remote sensing and spectral unmixing analysis for crop residue mapping. *Remote Sensing of Environment*, 114(10):2219–2228, 2010.

[97] M. S. Galloza, M. M. Crawford, and G. C. Heathman. Crop residue modeling and mapping using landsat, ali, hyperion and airborne remote sensing data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(2):446–456, 2013.

[98] C. S. T. Daughtry, P. C. Doraiswamy, E. R. Hunt Jr, A. J. Stern, J. E. McMurtrey III, and J. H. Prueger. Remote sensing of crop residue cover and soil tillage intensity. *Soil and Tillage Research*, 91(1):101–108, 2006.

[99] C. S. T. Daughtry and E. R. Hunt Jr. Mitigating the effects of soil and residue water contents on remotely sensed estimates of crop residue cover. *Remote Sensing of Environment*, 112(4):1647–1657, 2008.

[100] G. Serbin, E. R. Hunt, C. S. T. Daughtry, G. W. McCarty, and P. C. Doraiswamy. An improved aster index for remote sensing of crop residue. *Remote Sensing*, 1(4):971–991, 2009.

[101] J. Cihlar, L. St-Laurent, and J. A. Dyer. Relation between the normalized difference vegetation index and ecological variables. *Remote Sensing of Environment*, 35(2):279–298, 1991.

[102] F. A. Kruse, J. W. Boardman, and J. F. Huntington. Comparison of airborne hyperspectral data and eo-1 hyperion for mineral mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6):1388–1400, 2003.

[103] R. O. Green, B. E. Pavri, and T. G. Chrien. On-orbit radiometric and spectral calibration characteristics of eo-1 hyperion derived with an underflight of aviris and in situ measurements at salar de arizaro, argentina. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6):1194–1203, 2003.

[104] J. E. Morrison, C.-H. Huang, D. T. Lightle, and C. S. T. Daughtry. Residue measurement techniques. *Journal of Soil and Water Conservation*, 48(6):478–483, 1993.

[105] D. T. Booth, S. E. Cox, and R. D. Berryman. Point sampling digital imagery with samplepoint. *Environmental Monitoring and Assessment*, 123(1-3):97–108, 2006.

[106] M. S. Galloza, B. Engel, M. M. Crawford, G. C. Heathman, and J. R. Williams. Application of remote sensing observations as apex model input for estimating soil erosion. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Melbourne, Australia, 2013. IEEE.

VITA

VITA

Junhwa Chi received his bachelor degrees in Geoinformatic Engineering and Computer Science, the master degree in Geoinformatic Engineering at Inha University, Korea. He joined the Geomatics program in Civil Engineering of Purdue Unveristy in August 2009. He also enrolled in the interdisciplinary program of Computational Science and Engineering as a concentration in his PhD program. His research interests include manifold learning, spectral mixture analysis and related applications for hyperspectral remotely sensed data.